Markedness and Language Change

Empirical Approaches to Language Typology

32

Editors

Bernard Comrie
Matthew Dryer
Yaron Matras

Mouton de Gruyter
Berlin · New York

Markedness and Language Change

The Romani Sample

by
Viktor Elšík
Yaron Matras

Mouton de Gruyter
Berlin · New York

Mouton de Gruyter (formerly Mouton, The Hague)
is a Division of Walter de Gruyter GmbH & Co. KG, Berlin.

♾ Printed on acid-free paper which falls within the guidelines of the
ANSI to ensure permanence and durability.

Library of Congress Cataloging-in-Publication Data

Elšík, Viktor.
Markedness and language change : the Romani sample / by Viktor Elšík, Yaron Matras.
p. cm. − (Empirical approaches to language typology ; 32)
Includes bibliographical references and index.
ISBN-13: 978-3-11-018452-5 (alk. paper)
ISBN-10: 3-11-018452-4 (alk. paper)
1. Romani language − Markedness. 2. Romani language − Dialects. 3. Markedness (Linguistics) I. Matras, Yaron, 1963− II. Title. III. Series.
PK2897.E58 2006
491.4'97−dc22
2005036353

Bibliographic information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

ISBN-13: 978-3-11-018452-5
ISBN-10: 3-11-018452-4
ISSN 0933-761X

© Copyright 2006 by Walter de Gruyter GmbH & Co. KG, D-10785 Berlin.
All rights reserved, including those of translation into foreign languages. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording or any information storage and retrieval system, without permission in writing from the publisher.
Printed in Germany.

In memory of Milena Hübschmannová
1933–2005

Acknowledgements

Along with the Romani Morphosyntax (RMS) Database, this study is an outcome of a four-year research project devoted to the Morphosyntactic Typology of Romani Dialects. We are grateful to the Arts and Humanities Research Board for their support of the project (grants no. B/RE/AN4725/APN11878 and B/RG/AN4725/APN9447), and to the Open Society Institute's Roma Culture Initiative for a grant in support of the elicitation and processing of additional dialect material. A number of people were involved in the processing and archiving of data, and we are grateful to them: Charlotte Jones, Christa Schubert, Barbara Schrammel, Petra Cech, Irene Sechidou, Astrid Rader, Luzia Plansky, Katrin Hiietam, and Ioanna Sitaridou. A long list of fieldwork assistants participated in the data collection; we thank Milena Alinčová, Galina Aslanova, Irmela Bajramovska, Teofile Bogdanovich, Agnieszka Borda, Marie Bořkovcová, Olga Chashchikhina, Veliyana Chileva, Edouard Chilline, Kristina Dienstbierová, Petra Dobruská, Pilvi Duuna, Laszlo Foszto, Lýdia Gabčová, Amela Ismaili, Jelena Jovanović, Liljana Kovacheva, Radka Kováčová, Jana Kramářová, Martina Kubátová, Goran Lakatuš, Isabela Mihalache, Beata Oláh, Ana Oprisan, Indra Pařízková, Jelena Petrović, Helena Pirttisaari, Kristina Raducan, Petr Rubak, Veronica Schulman, Elvira Skenderovska, Eva Sobotka, Boyana Stanienova, Zuzana Strnadová, Sandra Sujová, Anton Tenser, Aspasia Theodosiou, Şirin Tufan, Mihaela Zătreanu, and Zuzana Znamenáčková. We also thank our colleagues at the Department of Linguistics at the University of Manchester, and especially John Payne, Nigel Vincent, and Kersti Börjars for their support of our work during the lifetime of the project, our friends and colleagues Peter Bakker, Victor Friedman, Dieter Halwachs, and the late Milena Hübschmannová for their encouragement of our research on Romani, and Bernard Comrie for providing comments on the manuscript and for accepting it for publication in the EALT series. Finally, we wish to thank Peter Kahrel for his support in preparing the manuscript for production.

Contents

Acknowledgements vii
Tables xvi
Illustrations xxi
Abbreviations xxii

Chapter 1. Introduction: Markedness and asymmetry in language 1

Chapter 2. The Markedness Hypothesis 7
2.1. Concepts of markedness 7
2.1.1. The structuralist/semiotic approach 7
2.1.2. The generative approach 9
2.1.3. The typological approach 10
2.1.4. The naturalness approach 13
2.2. Markedness criteria 15
2.2.1. Frequency 15
2.2.2. Conceptual complexity 17
2.2.3. Structural complexity 18
2.2.4. Distribution 19
2.2.5. System-dependent criteria 20
2.2.6. External criteria 22
2.3. Markedness and language change 22
2.3.1. The markedness reduction hypothesis 22
2.3.2. Type of change 23
2.3.3. Markedness and language contact 25

Chapter 3. Toward a communication-based model of asymmetry in language 28
3.1. Factors involved in the formation of asymmetry 28
3.2. Application of the model 32
3.3. Criteria for asymmetry 35
3.3.1. Complexity 36
3.3.2. Erosion 36
3.3.3. Differentiation 37
3.3.4. Extension 38
3.3.5. Extra-categorial distribution 39

3.3.6. Exposition. 40
3.3.7. Borrowing 41
3.3.8. Internal diversity 43
3.3.9. Criteria not included in this study 43
3.4. Factors motivating asymmetry. 44
3.4.1. Topical saliency 45
3.4.2. Transparency 45
3.4.3. Discourse accessibility 46
3.4.4. Egocentricity 46
3.4.5. Relevance 46
3.5. Concluding remarks 47

Chapter 4. The sample: Methodological considerations 48
4.1. Sampling in a typological context 48
4.2. Dialect sampling in Romani 50
4.2.1. The usefulness of dialect samples 50
4.2.2. The challenge of Romani. 52
4.2.3. Romani dialectology. 55
4.3. Putting typology to work in a dialect sample: The Romani Morphosyntactic Database (RMS) 58
4.3.1. The database tools 58
4.3.2. Function to form, form to function 60
4.3.3. Data collection procedures 63
4.4. Summary: Features and problems of the sample 65

Chapter 5. Early Romani. 68
5.1. Lexicon 69
5.2. The sound system. 70
5.3. Nominals 71
5.3.1. Case marking and declension classes 71
5.3.2. Adjectival modifiers 73
5.3.3. Demonstratives and related forms 75
5.3.4. Personal pronouns. 76
5.3.5. Interrogatives 77
5.3.6. Indefinites 77
5.4. Verbs 78
5.4.1. Valency and loan verb integration 78
5.4.2. Inflection classes. 80

5.4.3. Concord markers. 81
5.4.4. Tense, aspect and modality 82
5.5. Other categories. 83
5.5.1. Local adverbs 83
5.5.2. Prepositions 83
5.6. Syntax 84

Chapter 6. Number 85
6.1. Complexity 85
6.2. Erosion 88
6.3. Differentiation. 89
6.4. Extension 97
6.5. Extracategorial distribution. 99
6.6. Exposition. 100
6.7. Borrowing and internal diversity 101

Chapter 7. Person. 102
7.1. Complexity 103
7.2. Erosion 108
7.3. Differentiation. 112
7.4. Extension. 121
7.5. Extracategorial distribution 125
7.6. Exposition 133
7.7. Borrowing 134

Chapter 8. Gender 138
8.1. Complexity and erosion. 138
8.2. Differentiation. 140
8.3. Extension 141
8.4. Extracategorial distribution. 143
8.5. Internal diversity and borrowing 143

Chapter 9. Degree. 145
9.1. Complexity 146
9.2. Differentiation. 148
9.3. Borrowing and internal diversity 149
9.4. Extension 154

Chapter 10. Negation 156
10.1. Complexity 156
10.2. Differentiation 158
10.3. Extension 158
10.4. Internal diversity 159
10.5. Borrowing 160

Chapter 11. Cardinality 162
11.1. Complexity 162
11.2. Differentiation 163
11.3. Extension 164
11.4. Internal diversity 166
11.5. Borrowing 170

Chapter 12. Discreteness 173
12.1. Complexity 174
12.2. Erosion 181
12.3. Differentiation 182
12.4. Extension 183
12.5. Exposition 184
12.6. Internal diversity and borrowing 184
12.7. Linear order 186

Chapter 13. Tense, aspect, and mood 188
13.1. Complexity 189
13.2. Erosion 194
13.3. Differentiation 194
13.4. Extension 198
13.5. Extracategorial distribution 202
13.6. Borrowing 202

Chapter 14. Modality 203
14.1. Complexity 204
14.2. Differentiation 207
14.3. Linear order 208
14.4. Borrowing 209

Chapter 15. Transitivity 211
15.1. Complexity 211

15.2. Differentiation 213
15.3. Extension 215
15.4. Exposition 216
15.5. Internal diversity 216
15.6. Borrowing 216

Chapter 16. Case and case roles 218
16.1. Complexity 220
16.2. Erosion 223
16.3. Differentiation 224
16.4. Extension 229
16.5. Extracategorial distribution 230
16.6. Internal diversity 232
16.7. Borrowing 234

Chapter 17. Localisation 239
17.1. Complexity 243
17.2. Erosion 250
17.3. Differentiation 250
17.4. Extension 252
17.5. Extracategorial distribution 259
17.6. Internal diversity 264
17.7. Borrowing 267

Chapter 18. Orientation 271
18.1. Extension 273
18.2. Exposition 275
18.3. Internal diversity and borrowing 275
18.4. Complexity 279
18.5. Differentiation 280

Chapter 19. Indefiniteness 281
19.1. Complexity 283
19.2. Extension 284
19.3. Extracategorial distribution 285
19.4. Internal diversity 286
19.5. Borrowing 287

Chapter 20. Ontological category 295
20.1. Complexity 297
20.2. Erosion 302
20.3. Differentiation 303
20.4. Extension 304
20.5. Extracategorial distribution 305
20.6. Internal diversity 308
20.7. Borrowing 309

Chapter 21. Lexicality 312
21.1. Complexity 312
21.2. Differentiation 314
21.3. Extension 319
21.4. Borrowing 320

Chapter 22. Associativity 322

Chapter 23. Chronological compartmentalisation 324
23.1. Complexity 326
23.2. Differentiation 327
23.3. Extension 327
23.4. Exposition 332
23.5. Borrowing and diversity 332

Chapter 24. Criteria for asymmetry and their distribution across categories 334
24.1. Complexity 334
24.2. Erosion 336
24.3. Differentiation 337
24.4. Extension 339
24.5. Extracategorial distribution 340
24.6. Exposition 341
24.7. Internal diversity 342
24.8. Borrowing 343
24.9. Criteria relevance: Summary 344

Chapter 25. Patterns of asymmetry 347
25.1. The consistency of value ordering within categories 347
25.1.1. General considerations 347
25.1.2. Variation in linear order and polarity 352
25.2. Clusters of asymmetry criteria 355
25.2.1. Predictions of the Markedness Hypothesis and 'well-behaved categories' 355
25.2.2. Correlating criteria: Types of clusters 362
25.2.5. The position of borrowing 370

Chapter 26. Conceptual motivations for asymmetry 377
26.1. Iconic motivations for linear ordering 377
26.1.1. Quantity 377
26.1.2. Immediacy 379
26.1.3. Prominence 380
26.1.4. Truth and simplicity 381
26.1.5. Transparency 382
26.2. Global and local motivations 382
26.2.1. Exposition 383
26.2.2. Extension, distribution, and erosion 384
26.2.3. Borrowing 385
26.3. Conflicting hierarchies and conflict resolution 387
26.3.1. Conflict domains 388
26.3.2. Conflict categories 390
26.3.3. Conflict pairs 398
26.4. Motivations for asymmetry: Concluding remarks 404

Chapter 27. Concluding remarks 406

Appendix : Sample dialects 410
Notes 425
References 441
Index of authors 455
Index of Romani dialects 458
Index of geographical names 463
Index of subjects 465

Tables

Table 4.1. Middle inflections in selected dialects 51
Table 4.2. Reconstructed Early Romani determiners 54
Table 4.3. Inherited present-stem forms and their TAM function in some dialects 63
Table 4.4. Example of elicited sentences with tags (Polish Romani) 64
Table 4.5. Example of modal constructions with 'want' (Lithuanian Romani) 65
Table 4.6. Sample comparison of sentences in different dialects 66
Table 5.1. Early Romani consonant phonemes 70
Table 5.2. Early Romani nominal declension classes 72
Table 5.3. Early Romani Layer II case markers (Sg/Pl) 73
Table 5.4. Early Romani adjectival inflection 74
Table 5.5. Early Romani deictic and anaphoric expressions 75
Table 5.6. Early Romani first and second person pronouns 76
Table 5.7. Early Romani interrogatives 77
Table 5.8. Early Romani indefinites 77
Table 5.9. Early Romani perfective inflection classes 80
Table 5.10. Early Romani subject concord markers 81
Table 5.11. TAM categories in Early Romani 83
Table 6.1. Roots of personal pronouns 86
Table 6.2. Inflection of consonantal adjectives 87
Table 6.3. Intrusion in third person pluperfect inflections 88
Table 6.4. Differentiation asymmetries in the category of number 89
Table 6.5. Early Romani person–number suffixes 89
Table 6.6. Latvian Romani person–number suffixes 90
Table 6.8. Selected German/Austrian Sinti inflections 90
Table 6.7. East Ukrainian Romani person–number suffixes 90
Table 6.9. Hungarian Sinti non-remote non-perfective inflections 91
Table 6.11. Case homonymy in East Slovak Romani 92
Table 6.10. Manuš perfective inflections 92
Table 6.12. Case homonymy in Kosovo Bugurdži 93
Table 6.13. Case differentiation patterns in the first-person plural pronoun 94
Table 6.14. Inflection of xenoclitic adjectives in Early Romani 95
Table 6.15. Demonstrative inflection in Hungarian Lovari 95

Table 6.17. Non-perfective person–number suffixes in Early Romani 96
Table 6.16. Gender neutralisations in the third person pronouns 96
Table 6.18. Inflectional class differentiation in verb inflections 97
Table 6.19. Extension of singular demonstrative forms 98
Table 6.20. Singular-like oblique forms of selected pronouns 99
Table 7.1. Perfective inflections in selected dialects 104
Table 7.2. Perfective inflections in Abruzzian Romani 104
Table 7.3. Categorially determined distribution of indicative copula roots 107
Table 7.4. Erosion of the middle sequences 108
Table 7.5. Patterns of erosion in the middle sequences 109
Table 7.6. Variants of the remoteness suffix in the Northeastern dialects 111
Table 7.7. Differentiation asymmetries in the category of person 112
Table 7.8. Sofia Erli perfective inflections 113
Table 7.9. Rumelian Romani perfective inflections 114
Table 7.10. Slovak Romani (Zemplín) perfective inflections 114
Table 7.12. Finnish Romani preterite inflections 115
Table 7.11. Bougešťi perfective inflections 115
Table 7.14. Non-remote non-perfective inflections in Austrian Sinti 118
Table 7.13. Perfective inflections in earlier Finnish Romani 118
Table 7.16. Non-perfective inflections in Manuš 119
Table 7.15. Non-remote non-perfective inflections in Hameln Sinti 119
Table 7.17. Third-person pronouns in selected Balkan dialects 120
Table 7.18. Extensions in second-person plural and third-person plural perfective inflections: patterns 122
Table 7.19. Extensions in second-person plural and third-person plural perfective inflections: forms in selected dialects 123
Table 7.20. Person parallelisms in reflexive pronouns: forms 131
Table 7.21. Person parallelisms in reflexive pronouns: patterns 132
Table 7.22. Person–number inflections in Slovene Romani 135
Table 7.24. Preterite inflections in Kaspičan 136
Table 7.23. Perfective inflections containing the Turkic plural suffix *-Iz* . 136
Table 8.1. Demonstrative forms in Central Slovak Romani: the integration scenario 139
Table 8.2. Inflectional assimilation in demonstratives 141
Table 8.3. Early Romani nominative inflections of xenoclitic noun classes 143
Table 9.1. Types of the category of degree 145

Table 9.2. Borrowed degree markers according to their functions in the L2 and in Romani 150
Table 9.3. Distribution of degree markers borrowed from East South Slavic and Turkish 151
Table 9.4. Degree marking in dialects of Types V and VI 152
Table 11.1. Additive connectors in ten+unit numerals 166
Table 11.2. Construction types of ten numerals 168
Table 11.3. Distribution of Greek-derived ten numerals 171
Table 12.1. Complementiser differentiation in modal and manipulative clauses 178
Table 12.2. Complementiser differentiation in purpose clauses 180
Table 12.3. Demonstratives in selected dialects 183
Table 13.1. Indicative TAM values 188
Table 13.2. Mismatching TAM values in lexical verbs and in the copula 189
Table 13.3. Non-remote non-perfective forms 191
Table 13.4. Verb inflections in Roman 195
Table 13.5. Imperfective verb classes in Ajia Varvara 197
Table 13.6. Indicative and imperative negators 200
Table 13.7. Indicative, subjunctive, and imperative negators 201
Table 14.1. Distribution of tenses among types of conditional clauses . . . 205
Table 14.2. Complementiser te in modal complements with identical subject 206
Table 14.3. Borrowing of modal expressions 209
Table 14.4. Borrowing patterns of conditional particles 210
Table 15.1. Distribution of third-person singular active participles with different types of verbs 214
Table 16.1. Internal and external cases 218
Table 16.2. Differentiation asymmetries in the category of case 224
Table 16.3. Dative and genitive inflections in Kumanovo Gurbet 225
Table 16.4. Extensions of core and adverbial case roles into local and temporal domain 232
Table 16.5. External case suffixes: dialect forms 233
Table 17.1. Localisation values in Romani 240
Table 17.2. Early Romani local adpositions 241
Table 17.3. Early Romani local adverbs 242
Table 17.5. Orientation distinctions in core and axis local adverbs in Šóka Rumungro 251
Table 17.4. Orientation distinctions in case marking by localisation 251

Table 17.6. Adessive and inessive adpositions in the inessive localisation 253
Table 17.7. Extensions among core localisations 256
Table 18.1. Distribution of ablative forms of adpositions and adverbs . . . 274
Table 18.2. Local deictics in Rumungro 280
Table 19.1. Borrowed markers of free-choice indefiniteness 288
Table 19.2. Borrowed markers of negative indefiniteness 289
Table 19.3. Borrowed markers of specific indefiniteness 290
Table 19.4. Patterns of borrowing of indefiniteness markers. 290
Table 20.1. Ontological values in Romani interrogatives 295
Table 20.2. Determiner-base indefinites in selected dialects. 298
Table 20.3. Marking of nouns in adpositional case role ('behind') 301
Table 20.4. Patterns of erosion of the interrogative root s- 302
Table 20.5. Interrogatives as connectors in subordinate constructions . . . 308
Table 21.1. Intrusion *-in-* as stem extension 313
Table 21.2. Patterns of TAM differentiation in lexical verbs and in the copula 315
Table 21.3. Distribution of subject clitics 317
Table 22.1. Associative forms in Rumungro ('locksmith'). 322
Table 23.1. Domains and selected markers of chronological compartmentalisation 326
Table 23.2. Summary of extensions in chronological compartmentalisation 328
Table 24.1. Complexity asymmetries 335
Table 24.2. Erosion asymmetries 336
Table 24.3. Differentiation asymmetries 338
Table 24.4. Extension asymmetries 339
Table 24.5. Distribution asymmetries. 340
Table 24.6. Exposition asymmetries. 341
Table 24.7. Diversity asymmetries 342
Table 24.8. Borrowing asymmetries 343
Table 24.9. Summary of asymmetry criteria and their distribution across categories 345
Table 25.1. Presence of asymmetry hierarchies for a selection of categories. 348
Table 25.2. Presence of asymmetry hierarchies for additional categories. 350
Table 25.3. Presence of asymmetry hierarchies in ontological sub-categories. 351

Table 25.4. Linear order and relevant polarity for binary categories 353
Table 25.5. Summary of associations between hierarchies (Base Table) 357
Table 25.6. 'Well-behaved' categories 359
Table 25.7. Categories with 'default' values 361
Table 25.8. Links between Complexity and Differentiation 363
Table 25.9. Correlation of erosion and differentiation 364
Table 25.10. Correlation of erosion and complexity 365
Table 25.11. Correlation of erosion and extension 365
Table 25.12. Correlation of exposition and complexity 366
Table 25.13. Complexity and general susceptibility to change 367
Table 25.14. 'Markedness' and borrowing 371
Table 25.15. Borrowing and 'default' status 373
Table 25.16. Borrowing and internal diversity 374
Table 25.17. Polarity of borrowing 375
Table 26.1. Distribution of categories by iconicity principles 378
Table 26.2. Differentiation hierarchies in the category of internal case 392
Table 26.3. Complementiser *te* in selected modal complements with identical subject 398
Table I.1. British dialects of Romani 406
Table I.2. Northwestern dialects of Romani 407
Table I.3. Northeastern dialects of Romani 408
Table I.5. South Central dialects of Romani 409
Table I.4. North Central dialects of Romani 409
Table I.6. Slovene and Apennine dialects of Romani 410
Table I.7. South Balkan dialects of Romani 411
Table I.8. Balkan *zis*-dialects of Romani 412
Table I.9. North Vlax dialects of Romani 412
Table I.10. South Vlax dialects of Romani 413
Table I.11. Ukrainian dialects of Romani 414
Table I.12. Alphabetical list of Romani dialects 414
Table I.13. List of Romani dialects by map index number 416

Illustrations

Maps

Map 1. Locations of Romani dialects in southeastern and central Europe 417
Map 2. Location of Romani dialects outside southeastern and central Europe 419

Figures

Figure 3.1. The communication-based model of asymmetry in category paradigms 31
Figure 4.1. Sample entries for complementation 61
Figure 19.1. Semantic map of indefiniteness functions 281
Figure 19.2. Indefiniteness marking in Central Slovak Romani 282

Abbreviations

Gloss abbreviations

ABL	ablative
ACC	accusative
ART	definite article
ASS	associative
COMP	complementiser
DAT	dative
F	feminine
FUT	future
GEN	genitive (possessive)
IMP	imperative
IMPF	imperfect
INF	infinitive
IRR	irrealis, unreal conditional
ITR	intransitive
LIM	limitative particle
LOC	locative
M	masculine
MOD	modality expression
NEG	negator
NOM	nominative
NPFV	non-perfective
OBL	oblique
PFV	perfective
PL	plural
PLPF	pluperfect
PRES	present
PRET	preterite
REFL	reflexive
SG	singular
SOC	sociative (instrumental)
SUBJ	subjunctive
TR	transitive

Criteria and value abbreviations

1	first person
2	second person
3	third person
ABI	ability
ACC	accusative
ADJ	adjective
AFFIRM	affirmative
AKTIONS	aktionsart
ANIM	animate
ART	definite article
AUX	auxiliary
BOR	borrowing
COM	complexity
COMP	comparative
DET	determiner
DIF	differentiation
DIR	directive
DIS	extra-categorial distribution
DIV	cross-dialectal diversity
ERO	erosion
EXP	exposition
EXT	extension
FC	free-choice
FEM	feminine
FUT	future
GEN	genitive

HIGH	high quantity
IMP	imperative
INAB	inability
INANIM	inanimate
IND	indicative
IRREAL	irrealis
ITRANS	intransitive
LOW	low quantity
MAS	masculine
MODIF	modifier
NEC	necessity
NEG	negative
NEUTR	aktionsart-neutral
NOM	nominative
NON-AUX	non-auxiliary
NON-PERF	non-perfective
NON-REM	non-remote
OBL	oblique
PERF	perfective
PERIPH	peripheral relation
PERS	person
PL	plural
POS	positive
POSS	possessive
POT	potential
QUANT	quantity
REAL	realis
REM	remote
SEP	separative
SG	singular
SPEC	specifier
STA	stative
SUB	subjunctive
SUP	superlative
TRANS	transitive
UNIV	universal
VOL	volition

Chapter 1
Introduction: Markedness and asymmetry in language

Symmetry is arguably one of the most deeply rooted principles of human cultural aesthetics. It derives at least in part from an appreciation of symmetry in nature and a need to replicate the external symmetry of the human body. On the body map, symmetry is manifested by the position and shape of organs that have equivalent functions. We tend to replicate this bodily symmetry by shaping items that are functionally equivalent in a matching fashion, and by positioning them in matching locations relative to a stable point of reference (often determined by the position of our own body and the direction it faces). The aesthetic effect of symmetry is the satisfaction we get from recognising an image of our bodies, and of other living creatures, in the well-formedness of symmetrical artefacts.

This emotional satisfaction is often coupled with a realisation that symmetry can be functional and effective. The need to balance loads against the gravitational force is a natural, utilitarian trigger for symmetry. It finds its expression in anything from the design of a trash bin, to the wheels of a car or the wings of an aeroplane. We might therefore assume that much of what we could call cultural symmetry is driven by a combination of motivations: an anticipation and appreciation of symmetry as both naturally well-formed, like the human body, and as effective and functional, like a balanced object whose course or position are steady and predictable. Take the symbolic presentation of the Ten Commandments as an example: They are portrayed as written on two stone tablets, as described in Exodus, while on ornaments and other artistic depictions the longer commands are abbreviated so that each tablet contains five of the ten. Aesthetic symmetry is thus complemented by the efficiency-oriented symmetry of a balance of loads.

But as McManus (2003) points out, symmetry is not quite the dominant pattern of organisation that we might expect, in either tools or organisms. For a start, body symmetry is only external, and even external bodily symmetry is rarely consistent, as can be seen in differently shaped teeth, feet, breasts, nostrils, eyes and more. The more outstanding asymmetries in the human body are of course the position of the heart and other internal organs, and lateralisation

in the brain. The latter is responsible for asymmetry in handedness, which is a clearly visible part of human experience as it conditions the patterns of basic physical activities. It also means that the neurofunctional organisation of language and other expressions of cognition and consciousness in the brain is asymmetrical.

If the shaping of linguistic structures were an entirely conscious and aesthetically oriented procedure, we might expect symmetry to play some role, though it is likely that asymmetry too would be tolerated. For instance, if a language were to express the subject-concord marker of the first person by means of a clearly analysable monosyllabic suffix that immediately follows the verbal root, then we might expect the subject-concord marker of the second person to appear in the same position, and to take a similar shape, i.e. also that of an analysable suffix. The need for some regularity in form–function correlation is clearly observable in language, as in the well-noted example from child language acquisition, where *play*/*played* is matched by *go*/*goed*. Beyond the aesthetic aspects, regularity is useful, as it enables us to predict the shape and position of structure components that carry similar functions. However, functionality also constrains regularity: in our example of concord suffixes, it may be aesthetic (that is, answering the expectation of well-formedness and the need for predictability) for both suffixes to occupy a similar position in the word, but it would obviously be dysfunctional for them to have identical shape. Like the right hand and the left hand, the two suffixes will play different roles within the paradigm to which they both belong.

The essence of linguistic categorisation is to postulate relations among forms that perform complementary, but separate functions (some might say, similar functions but separate *meaning*; we consider the conveying of *meaning* to be part of the function of a linguistic structure). The traditional framework for categorising functions is the structuralist notion of the paradigm, where the functional relationship is essentially determined by a substitution test. If one form can replace another in the same environment, then the two are viewed as members of the same paradigm, and so as functionally related. The difference between them is internal, but crucial, since it allows picking one over the other in pursuit of a particular communicative objective.

Our hypothetical example of symmetry in the shape and position of concord affixes is not necessarily the general rule in language, but it is not an exception, either. The fact that paradigms tend to have concrete structural manifestations is evidence that individual linguistic structures are mentally assigned to paradigms on the basis of similarities in their function and the type of their relationships to other linguistic structures. At the same time, despite

being categorised as similar, the raison d'être of each and every member of a paradigm is that it differs from other members in its internal function.

Consequently, we observe a tension between two tendencies in language. The first is to treat members or *values* of the same paradigmatic category as status-equivalent in terms of their position and shape; this is the tendency toward symmetry among members of the paradigm, which is reminiscent of the external symmetry among some body parts. The second tendency is to give expression to distinctive features of a member or value within a paradigm, by structuring it differently from the other members, thereby disturbing the overall symmetry of the paradigm. In the course of the next chapter, we discuss the criteria by which we can recognise the absence of paradigmatic symmetry. The working assumption in the study of asymmetry is that differences among members of a paradigm are often not accidental. Rather, they mirror categorisations of knowledge and information as salient or relevant in different ways or to different degrees. These are cognitive processes that complement the cognitive demand for regularity and predictability in the organisation of linguistic forms performing similar functions.

Asymmetries in paradigms have come to be associated with the notion of 'markedness'. The concept assumes that the structural relationship between two poles on a paradigm is predictable to some extent. It also assumes that one of the poles, the 'marked' counterpart, will consistently display properties that the other, 'unmarked' pole, lacks. A good example of a metaphorical, and yet still linguistic-methodological use of these notions is the way they have been applied to entire linguistic repertoires in Myers-Scotton's (1993) 'Markedness Model of Codeswitching'. There, a 'marked' choice of language is an unexpected choice, one which goes against the conventions assigned by the bilingual speech community to the specific type of situation, and so a choice that is used to challenge an aspect of the speech context or to contrast it with contextual expectations. This use of the notion of 'markedness' derives from the original Jakobsonian and Trubetzkoyan view of markedness as the presence of a feature, as opposed to the absence of this feature in the counterpart 'unmarked' structure. This dualism approach to markedness is still popular in structuralist discussions of variation, where one variant, for instance a particular word-order pattern in a language with some word order flexibility, is regarded as the default choice, while the other is seen as exceptional and, though still grammatical, as a statement that challenges expectations.

Against this background, it is sometimes expected that any two counterpart structures which co-exist in a paradigm and form a binary opposition of

some kind might be divided hierarchically into a 'marked' and an 'unmarked' member. Moreover, it is sometimes expected that a theory of language might be able to predict 'markedness', interpreted in this way, for certain pairs of structures, based on the nature of the internal (meaning- or function-based) opposition between them. One member of the pair, it is argued, will always be 'marked', while the other will always be 'unmarked'. We shall refer to this view as the 'Markedness Hypothesis'.

We argue against this notion in this book. We return to our assumption that language reflects tension between two competing tendencies – one toward regularity, the other toward hierarchisation (Croft 2003 speaks of a tension between the economy of language, and its iconicity). Both reflect cognitive universals of communication, which rely both on predictability and regularity, and on prioritising information in accordance with hearer-sided expectations (see Givón 1984, 1990). The task of linguistic theory is, in our view, to describe how these competing tendencies are responsible for shaping linguistic structures. This involves an interplay, at the local level, of several factors: the structure and its function in communication, the member and the value it represents both within the paradigm and in a more universal conceptual framework, the nature of the specific process involved in shaping the structure, and the motivation to apply this process to the paradigm or to parts of it. This complex interplay of factors is never pre-determined, since different combinations of factors will render different results. In this respect, we take an 'unbiased' view of what 'markedness' is, and whether or not one structure is 'marked' compared with another. Rather, our interest is in exploring patterns in the outcomes of different factor combinations at the local levels.

Our view of asymmetry as a whole is not pre-determined, either. Part of our agenda will be to investigate whether, in what might be considered an inflectional language, with predominantly finite structures and some flexibility of word order, any one of the two patterns – symmetry or asymmetry – may be considered quantitatively more dominant. If we choose nonetheless to single out asymmetries in the discussion of our data, it is because symmetry seems more straightforward: It can be explained, in formal terms, at the level of the paradigm, and in functional terms, through the need for regularity in the position and shape of structures that perform similar linear operations in the organisation of information in the utterance. The forces that trigger asymmetry are, by contrast, much more opaque. They compete at various local levels against the seemingly overwhelming and ever-present power of the quest for symmetry. We therefore devote our investigation to these local manifestations of asymmetry, and the driving forces behind them.

To our knowledge, no attempt has yet been made to provide a systematic and exhaustive account of structural asymmetries in any individual language. Nor are we aware of a study that analyses the role of asymmetry in language change by taking into consideration closely related varieties and a stable time depth factor. Our aim is to use synchronic variation among our sample varieties in order to identify and describe the role which asymmetry plays in language change.

Although there is no historical documentation of the development of Romani, we can assume that present-day varieties of Romani descend from a common ancestor, which we shall call Early Romani, and which was spoken from the tenth or eleventh century and up until the late fourteenth century in Byzantium. The language of a socially marginalised group of itinerant immigrants originally from India, Early Romani was not written, and it enjoyed no institutional support. The nature of relations between the Romani community, who maintained a service-economy, and the majority, who were their clients, made bilingualism absolutely vital to the Rom. Consequently, their language absorbed influences from the surrounding languages. This remained the pattern of language use long after the decline of the Byzantine Empire and the dispersion of Romani-speaking populations throughout Europe. Groups became isolated from one another, and their languages diverged not just internally, but became exposed to the massive influence of contact languages as different as Basque, Turkish, German, Finnish, Italian, Polish, Rumanian, and Hungarian.

The Romani sample provides us therefore with useful conditions to test the role of asymmetry in shaping linguistic structures: We have (a) a measurable time depth of divergence, of up to seven centuries; (b) despite the lack of documentation, a reconstructable point of departure (owing to dialect comparison and reliance on documentation of related, medieval Indo-Aryan languages); (c) dispersion and isolation of the dialects, so that one can speak of a sample of closely related languages; (d) strictly oral traditions and no institutional use of language, and so the most natural patterns of development of speech; (e) continuous bilingualism, involving a variety of different languages but under comparable sociolinguistic conditions, and a tendency to absorb massive influences from the respective contact languages. Our agenda is to make use of the opportunities offered by the Romani sample in order to investigate (a) asymmetries in structural representation among individual members or values of functional categories; (b) the ways in which processes of change and structural formation may affect different members or values of the same category; (c) trends in the way different processes of change and structural formation target clusters of values across categories; and (d) the conceptual

background which motivates unbalanced changes, resulting in local expressions of asymmetry.

The first part of the book is devoted to our research questions and research tools. In Chapters 2 and 3 we survey the context of research on markedness, and define our approach to the topic known as 'markedness' and our agenda for the present investigation in more detail. Chapter 4 outlines our sampling and evaluation methodology. We open with a survey of the present research context on Romani, discuss sampling methods in typology, and outline the Romani Morpho-Syntax (RMS) database, which has served as the principal tool for the organisation of our data. Chapter 5 offers a reconstruction of Early Romani, the common point of departure for the Romani varieties, against which changes are assessed.

The second part of the book, Chapters 6 to 23, provides a survey, by category, of asymmetries in the Romani sample. We explore how different processes of change and structural criteria, such as erosion, complexity, borrowing, or differentiation, affect individual values (or members) of a category (such as first, second and third persons), taking into consideration all relevant structural manifestations and distributions of that category (for instance, the category Person appears in pronouns, as well as in concord marking on the verb).

The third part of the book is devoted to an evaluation of these data. Chapters 24 and 25 review consistencies and inconsistencies in the ordering of category values, and the ways in which, for individual categories (e.g. Person, Negation, or Location), different criteria of asymmetry may cluster into general tendencies (for example, whether more complex is also more prone to borrowing, etc.). Chapter 26 concludes with an attempt to relate these tendencies to conceptual motivations to treat values in different ways. Here, we return to the hypothesis that language shows competing tendencies toward symmetry among functionally related structures on the one hand, and toward a ranking of values based on their distinctive internal function, on the other. We assume that such ranking of values, which results in structural asymmetry, is anchored in cognitive universals of mapping experience onto the organisation of communication. We conclude with a discussion of the functionality of asymmetry in language change (Chapter 27).

Chapter 2
The Markedness Hypothesis

In this chapter we survey several approaches to markedness, criteria that have been used to identify markedness patterns, and claims about the connection between markedness and language change. Our own approach to the topic known as 'markedness' is outlined in Chapter 3.

2.1. Concepts of markedness

2.1.1. The structuralist/semiotic approach

The structuralist/semiotic approach to markedness builds on the original concept of markedness as developed by two prominent Prague School structuralists, Trubetzkoy and Jakobson. An overview of the semiotic markedness paradigm is given in Battistella (1996: 19–49). Trubetzkoy (1939) introduced the concept of markedness in the context of his research on phonological correlations. He viewed markedness relations as oppositions between the presence of some phonological feature and its absence in the consciousness of speakers. Later, as he restricted the applicability of markedness to neutralisable phonological oppositions, **neutralisation** became the defining criterion of markedness.

Jakobson, in his long-lasting and ever developing work on the concept (esp. Jakobson 1932, 1936, 1939, 1941, 1957), unfolded the potential of markedness in various directions and extended its application beyond phonology, to other linguistic levels and semiotic domains. Realising that markedness relations may be imposed not only on phonological oppositions but also on semantic categories in grammar and culture, he developed a global view of markedness as a general **value relation** between oppositions, which is applicable on different levels of analysis.

Although the nuances kept changing over time, Jakobsonian markedness may be defined as an asymmetrical relation between signalisation of a certain property (in the marked member of an opposition) and **non-signalisation** of that property (in the unmarked member of an opposition). Thus markedness was viewed as a **binary** relation, or at least as decomposable into binary

relations. Two levels of unmarkedness were distinguished: on the level of general meaning, the unmarked member of an opposition implies no statement about the relevant property of the marked member, while on the level of specific (or nuclear) meaning, the unmarked member indicates the opposite value of the relevant property. Andersen (1989, 2001) stresses the inclusive character of the markedness relation. The role of **general meaning**, or semantic invariance, in markedness has been taken forward especially by van Schooneveld (1978) and Andrews (1990).

In the course of Jakobson's work, understanding of the character of the relevant semiotic property of the marked member changed from that of a substantive property of the objective reality to that of a language-specific categorial **value**, a development that was in congruence with the work on markedness by Hjelmslev (1935) and later structuralists. Thus, for Andersen (1989), the markedness relation is in part independent of linguistic substance and should be defined primarily as conceptual. The value-oriented approach made it possible to extend the application of markedness to uses of categories and grammatical constructions.

In his late work (Jakobson and Waugh 1979), Jakobson contrasted phonological markedness, defined in terms of acoustic and typological properties, with semantic markedness, defined in terms of asymmetrical value relationships, although he saw an "intrinsic commonality" between the two concepts. According to Waugh (1982), markedness is a general structural principle underlying any system of oppositions, where the marked member shows **constraining**, focusing characteristic and conveys a more narrowly specified and delimited item. Similarly, Shapiro (1972) suggests that the marked member of an opposition has a narrower referential scope and greater conceptual complexity. Chvany (1985) observes that the key word uniting all kinds of markedness is informativeness.

Jakobson observed a number of correlations of markedness, which have been later worked out into markedness diagnostics by linguists of various approaches, and will be addressed in Section 2.2. First, he (1932) noticed that unmarked values tend to be represented by zero forms. This observation was elaborated on in his typology of zero signs (1939), where however zero forms (e.g. zero sound quality, zero phoneme, or zero desinence) were discussed alongside neutral or default distributions and functions: e.g. his zero opposition (i.e. neutralisation), zero morphological function (i.e. grammatical homonymy), zero meaning, zero (i.e. dominant) word order, or zero expressivity (i.e. stylistic neutrality). In his work on **iconicity** (1957, 1958), Jakobson suggested that there is a correspondence between semantic features and their phonologic-

al expression: not only do marked values tend to be encoded by overt markers, but also semantically proximate values of a category tend to be expressed by phonologically or phonotactically similar markers (so-called partial syncretism). He was nevertheless aware of the possibility of markedness conflicts.

Second, incorporating Brøndal's (1940) principle of compensation into his agenda, Jakobson (1936, 1939) demonstrated that marked values tend to show less formal differentiation than unmarked values. He also made a start on the issue of markedness **reversals** (i.e. reversals of markedness values in marked contexts), which was later developed especially in the works of Andersen and Shapiro. Andersen and Shapiro also elaborated on the related notions of markedness **assimilation** (i.e. assimilation of markedness of form to markedness of the context) and its opposite, markedness complementarity. Thus, markedness patterns in this current of neostructuralist theory – which Battistella (1996: 35–40) subsumes under the heading of *value iconism* and contrasts with the research on *semantic invariance* – are highly context-dependent. Battistella critisises the unclear basis for identification of markedness values and for motivation for reversals etc. in the value iconism approach.

Third, in an important study on first language acquisition and aphasia (1941), Jakobson discovered **extralinguistic** correlates of markedness in a universal hierarchy of phonological features. He collected data showing that marked features are more difficult for children to learn and easier for aphatics to lose, and he observed that the acquisition hierarchy has a clear correspondence in cross-linguistic distribution of the features, and may be formulated in a set of implicational universals.

2.1.2. The generative approach

The approach to markedness adopted by generative theories of language originally drew on Jakobsonian ideas about the concept, but it has been developing in its own direction since Chomsky and Halle (1968), where the concept of markedness entered the generative scene. Our brief overview of the approach is based on an extensive discussion in Battistella (1996: 73–123). He notes that Chomsky's view of markedness shows a remarkable flexibility, that the concept has not been developed in a systematic manner, and that "it is difficult to talk about there being any elaborately worked out theory of markedness in Chomsky's work" (p. 92).

Nevertheless, Battistella identifies two key ideas in the generative approach to markedness. First, markedness is conceived of as encoding a **preference**

structure or default structure for language acquisition (which is an idea shared with the semiotic and the naturalness approaches; see Section 2.1.4 for the latter). And, second, markedness is viewed as reflecting the **cost** of particular analytic options. The two ideas are interconnected in the generative approach, inasmuch a formal theory is required to exhibit explanatory adequacy with regard to language acquisition.

The concept of markedness was initially explored as part of the evaluation metric, which is a theory-internal construct that enables the linguist to select the most highly valued grammar. Marked and unmarked came to be understood as costly and cheap, respectively, in terms of the evaluation metric. Thus, markedness in the generative approach applied not only to linguistic elements, but especially to **descriptive formalisations** (rules, conditions, rule orders, transformations etc.).

In the late 1970s and in the 1980s, markedness began to be treated as part of a theory of core grammar. Core grammar consisted of a few parameters that were to be fixed during acquisition of an actual language, and was opposed to grammatical periphery that added rules to, and relaxed rules of, the core. The concept of markedness was applied doubly in this framework. First, the whole core grammar was considered to be unmarked as against the marked periphery. Markedness of a construction was determined by its regularity, stability, and centrality to the core of a particular language, as well as by cross-linguistic generalisations about construction types. Second, markedness also applied to parameter values within the core and within the periphery. Thus, markedness was also viewed as a preference structure within the two components of grammar.

2.1.3. The typological approach

The defining feature of the typological approach to markedness is the employment of **cross-linguistic** evidence. What Battistella (1996) calls the consistency problem in markedness theory – namely whether markedness is viewed as a universal or a language-particular concept – is resolved in favour of the former option. Although the research of cross-linguistic aspects of markedness originates in the works of the Prague School (Trubetzkoy 1939; Jakobson 1941, 1958), the typological approach has been fully developed only in the theoretical works of functional typologists (Greenberg 1966; Croft 1990, 2003; Givón 1990).

Croft (1990) introduced the term typological markedness, in order to distinguish it from distinct concepts of markedness in other schools of thought.

He defines typological markedness as a network of relationships among cross-linguistic asymmetrical patterns in grammar. Some points of this definition are briefly discussed below.

The view that markedness represents the fact of **asymmetrical** properties of otherwise equal linguistic elements is fully compatible with the other approaches to markedness. According to Croft (2003), typological markedness only concerns encoding of **function** in grammatical form, and thus asymmetrical patterns in word-order and phonology differ significantly from typological markedness in morphosyntax. The exclusion of phonological and word-order asymmetries is an important difference relative to the position of Greenberg (1966) and Croft (1990). Thus, in Croft's (2003) view, typological markedness is a property of **conceptual** categories. More specifically, it is a relationship between **paradigmatically** related values of conceptual categories.

Markedness is instantiated by cross-linguistic patterns that may be formulated as **implicational** universals, i.e. as constraints on logically possible combinations of linguistic properties. For example, if the marked value, such as the plural, is expressed by the absence of a morpheme, then so is the unmarked value, such as the singular (cf. Croft 2003: 89). The observation that the implicatum of an implicational relation that contains paradigmatically related values is usually the unmarked value goes back to Jakobson (1958), who applied the notion to explain some aspects of first language acquisition and aphasic dissolution. Eckman's (1977) Markedness Differential Hypothesis (see Section 2.3.2) makes use of the implicational aspect of markedness in predicting which structures will and will not be difficult in second-language acquisition. The implicational aspect of markedness has also been used in predicting possible and impossible diachronic changes. For example, the synchronic generalisation "if the plural is expressed by the absence of a morpheme, then so is the singular" involves the diachronic prediction that a zero plural will not develop unless a zero singular develops as well.

Markedness in the typological approach is viewed as a **network** of relationships. It subsumes a set of logically independent general patterns which, ideally, all select the same value as the unmarked value (Croft 1990, 2003). These general patterns are the **criteria** or diagnostics of markedness. In Battistella's (1996:53) words, the typological markedness theory is "a theory of correlations". Greenberg (1966) surveyed a number of correlates of markedness proposed by the structuralists in a cross-linguistic perspective. Croft (1990, 2003) reclassified Greenberg's criteria of markedness into three broad types: structural criteria, behavioural criteria, and token frequency criteria. Structural and behavioural criteria concern language structure, while frequency criteria

concern language use. Unlike structural criteria, behavioural and frequency criteria are universally applicable in morphosyntax, and thus they are considered to be more powerful diagnostics of markedness.

In Croft's (2003) approach to typological markedness as concerning only areas of language that involve form–function mapping, some of the criteria of morphosyntactic markedness are not applicable for asymmetrical patterns in word-order and phonology, and vice versa. Moreover, some widely used markedness criteria (e.g. neutralisation) are viewed as **invalid** for typological markedness on the grounds that there is no cross-linguistic consistency as to which value is unmarked (e.g. there is no consistent cross-linguistic pattern of neutral contexts that can be linked to other criteria of typological markedness). An important characteristic of Greenberg's (1966) and Croft's (1990, 2003) approach to markedness is that intrinsic semantic properties (such as conceptual complexity or informativeness) are not employed as criteria of markedness; and since the network of markedness criteria defines markedness, semantic properties are not part of the definition (cf. Battistella 1996:51). On the other hand, Givón (1990) recognises cognitive complexity as a markedness criterion, alongside formal complexity and frequency.

Markedness in the typological approach involves **relative** quantitative asymmetries between the formal expression of values. The view of markedness as a relative notion is shared with the naturalness school (see Section 2.1.4), and is opposed to the notion of markedness in the semiotic approach, which attempts to reduce markedness relations to binary oppositions or systems thereof. In the typological approach, a category value is more or less marked, rather than singly or doubly marked as in the semiotic approach. According to Croft (2003), the fact that neutralisation is not a relative concept explains why it is an invalid criterion of typological markedness. The recognition of the relative (or gradual or scalar) character of markedness enables one to draw in some fundamental concepts of linguistic typology into the markedness agenda. Apart from simple implicational universals, **hierarchies** and **prototypes** may be also viewed as markedness patterns (see Croft 2003 for a detailed discussion).

The phenomenon of **local markedness**, an analogue to markedness reversals in the semiotic approach, may be conceived of as a sort of prototype (cf. Croft 2003: 165). For example, the imperative is prototypically associated with the addressee, and so the second person is unmarked in imperative constructions, although it is not unmarked in other contexts (cf. Greenberg 1966: 44). Tiersma (1982) views local markedness as principled and explainable exceptions to general markedness patterns. The domain of local markedness

is definable on the basis of **real-world** (biological and/or cultural) considerations. Local markedness is said to apply to sets of semantically similar lexical items, while general markedness refers to categories. The use of words in a particular societal context is more important for local markedness than their lexical properties. Local markedness is claimed to be a matter of degree, so that certain domains may show competition of patterns of general and local markedness. Local markedness is in effect viewed as an implicative relation: the absence of evidence in a language for local unmarkedness within a relevant domain does not falsify the concept, as general markedness may be invoked; on the other hand, "evidence that certain words show the effects of local markedness without fitting the semantic or real-world criteria which have been associated with it" (Tiersma 1982: 847) would be considered as counterevidence.

Comrie (1986: 85) attempts to account for markedness in terms of "independently verifiable properties of people, the world, or people's conception of the world". Markedness is viewed as explainable in terms of human interaction with other humans and with the world, and not as an accidentally inherited or a purely formal property of language. There is a correlation between linguistic (formal) markedness and **situational markedness**, i.e. conceptualisation of extralinguistic situations. An unmarked situation is the one that is natural, expected, and has some real-world likelihood.

2.1.4. The naturalness approach

The naturalness approach to markedness has been developed in the school of Natural Morphology, which arose in Austria and Germany in the mid 1970s. The proponents of Natural Morphology (Dressler et al. 1987) characterise their approach as semiotic and, at the same time, functional. The school has been inspired, among others, by the Praguean markedness theory. On the other hand, Natural Morphologists share a number of theoretical viewpoints with the functionalist paradigm in linguistics, and especially with function-oriented typology (such as the assumption of relative character of markedness, the assumption of prototypes, and the reliance on extralinguistic motivations and extralinguistic evidence).

Natural Morphology is developed as a theory of **naturalness** in morphology. The theory recognises several levels of linguistic analysis, which are modelled by corresponding subtheories. The markedness theory models human language faculty and linguistic universals; the typological theory concerns

language types; and the theory of system congruity models language-specific competence. Language-specific norms and performance should be modelled by sociolinguistic and psycholinguistic ramifiactions of Natural Morphology. The subtheories are viewed as special cases of preference theory. While the theory of system congruity is based on 'normality' interpretation, the markedness theory in Natural Morphology is based on biological and socio-communicational interpretation. Natural Morphologists consider the generative and typological approaches to markedness as based on 'normality' interpretation.

Distinct **preferences** hold for different parameters (or scales) of naturalness (such as the morphological naturalness principles of constructional iconicity, uniformity, transparency, system congruity, or stability of morphological classes). The markedness theory assigns universal preferences to grammatical techniques of each naturalness parameter. Language types (whose inventory is taken over from Skalička's Prague School typology; e.g. Skalička 1979) are said to be constituted by specific constellations of selections from the various naturalness parameters: a language type sacrifices the naturalness of some parameters for the sake of greater naturalness of other parameters. Individual languages represent more or less perfect realisations of language types by exhibiting so-called system defining structural properties; compliance of a construction with system defining properties is the construction's degree of system congruity.

One of the aims of the theory of Natural Morphology is to establish types of possible **conflicts** between naturalness principles. Two sorts of conflicts are recognised: (a) conflicts within morphology between (morphological) naturalness principles (e.g. the conflicts in type-specific settings of different naturalness parameters, and the conflicts between universal and language-specific naturalness), and (b) conflicts between different components of the language system. Conflicts of the latter type are viewed as necessary, due to the relative autonomy of each component and its tendency to follow its own principles of naturalness. As the ultimate source of naturalness conflicts are the diverging functions of language components, Natural Morphology is bound to rely on functional explanations.

Generally, naturalness is claimed to have **extralinguistic** foundations, which determine/prohibit or favour/disfavour linguistic structures. In other words, extralinguistic facts constrain the possibilities, and assign the preferences, of the universal language faculty. However, although Natural Morphology must refer to extralinguistic facts, linguistic facts are not reducible to extralinguistic facts. Extralinguistic factors are divided into (a) neurobiological ones (e.g. limitations on perception and receptive processing [e.g. the prin-

ciple of ground and figure favours processing of salient, contrastive stimuli], limitations on selective attention, limitations of memory, storage and retrieval of information); and (b) socio-communicative ones (e.g. the speaker's empathy with the hearer's receptive role – presupposed by the trade-off between production effort and reception ease – depends on speech situation, the social roles of interlocutors, expectations etc.). Neurobiological and socio-communicative factors may interact (e.g. the definition of prototypical speaker is based on the pragmatic structure of a speech situation as well as on human sensory system). In short, universal naturalness corresponds to the ease "for the human brain" (Dressler et al. 1987: 11).

Due to the superior communicative and cognitive function of verbal signs (whose system constitutes the language) over non-verbal signs, Natural Morphology is explicitly sought to be framed in **semiotics**, "the most promising candidate to supply a meta-theory for [Natural Morphology]" (Dressler et al. 1987: 15). The questions asked in Natural Morphology analysis include semiotic adequacy of a linguistic sign (e.g. in terms of its size, position, and redundancy). In the evaluation of naturalness of a morphological technique, semiotic principles may conspire or come into conflict. While all linguistic signs are symbolic to some extent, icons are claimed to be the most natural signs, as they precede symbols both in phylogenesis and ontogenesis. Unlike images (icons of direct quality) and metaphors (icons of parallelism and partial similarity), diagrams (relational icons) represent the dominant semiotic interest of Natural Morphology. The term iconicity is used for relations between symbolic-diagrammatic signs.

2.2. Markedness criteria

2.2.1. Frequency

The criterion of frequency has been employed, and has occupied a central position, in the typological approach to markedness. The criterion concerns frequency, in texts, of tokens of linguistic elements of various sorts. Hence the more precise labels *token frequency* or *text* (*discourse*) *frequency*.[1] Token frequency is distinct from type frequency (e.g. of a stem within a morphological paradigm).

Greenberg (1966) has demonstrated that elements of unmarked values are more frequent that those of marked values. Croft (2003: 111) formulates the (intralinguistic) criterion of frequency as follows: "if tokens of a typologically

marked value occur at a certain frequency in a given text sample, then tokens of the unmarked value will occur at least as frequently in the text sample". Based on counts of linguistic elements, the criterion is relative and universally applicable. It is stressed that counts for typological markedness must count conceptual values, not linguistic forms. For example, one should count nouns with singular reference and not just nouns in the singular, since nouns in the singular can also be used for plural reference.

Following Greenberg, Croft (1990: 85) offers also a cross-linguistic version of the frequency criterion, where the domain of frequency counts is a language sample rather than a text sample, and the items of frequency counts are languages rather than tokens of linguistic elements. The cross-linguistic criterion is abandoned in Croft (2003). Meier (1999) questions the universal validity of the frequency criterion, showing that presumably marked values (e.g. glottality in stop consonants) have a very different distribution in different languages. She proposes to consider the frequency criterion to be language specific only.

According to Croft (1990: 84, 160), the frequency criterion shows a direct connection between properties of language structure and properties of language use. Token frequency is an extragrammatical factor that, however, imposes constraints on possible grammars. Token frequency of linguistic elements has been employed not only as a criterion of markedness but also used to explain or motivate the phenomenon. Greenberg (1966: 65–69) considered frequency to be the primary determining factor of markedness in morphosyntax. Croft (1990: 156–158, 2003) elaborates on the role of frequency as a causal source of other markedness criteria via economy. Following Haiman's (1985) ideas, he discusses economical motivation of structural and behavioural criteria of markedness: the most frequent (unmarked) forms are likely to be physically shortened and, ideally, zero coded; and the least frequent forms are more likely to be regularised or disappear, and hence greater inflectional potential and irregularities tend to be preserved only in the less frequent (less marked) forms.

As Battistella (1996) points out, if markedness was fully explainable by, and hence reducible to, a single criterion, then there would be no need for an independent concept of markedness. Indeed, this is what Haspelmath (2002: 237–238) suggests in his review of frequency effects in morphology: "asymmetries in the behaviour of inflectional categories that belong to the same inflectional dimension […] can be straightforwadly described and explained in terms of frequency differences, so that we do not need to make reference to abstract 'markedness'".

Nevertheless, Greenberg's and Croft's position is not that straightforward. In other passages of his book, Greenberg (1966) suggests that frequency itself is actually a symptom of, rather than the ultimate motivation for, markedness. The impossibility of explaining the criterion of typological distribution by frequency leads Croft (1990: 159) to conclude that markedness is more than just a manifestation of economical motivation. He also acknowledges the necessity to look for further causes of the frequency of certain grammatical values in speech. Text frequency is said to reflect a combination of real world facts and human choices in talking about the real world, i.e. characteristics of human cognition and communicative choices. The factor determining token frequency of semantic categories is their prominence to humans. The prominence factor has been also termed salience or expectedness (cf. Comrie 1986).

2.2.2. Conceptual complexity

In the structuralist/semiotic approach, conceptual complexity is the defining property of semantic markedness. Jakobsonian markedness (see Section 2.1.1) is, on the level of general meaning, defined as an asymmetrical relation between signalisation and non-signalisation of a certain property. Thus, the marked member of an opposition is by definition more semantically complex (more informative, more focused, showing narrower referential scope) than the unmarked member. Other markedness criteria are at best correlates or diagnostics of conceptual complexity. Positions of individual researchers in the structuralist/semiotic school of thought differ as to whether conceptual complexity requires substantive or contextual definitions.

Conceptual complexity has differing theoretical status in different typological and functional approaches. In Greenberg's (1966) and Croft's (1990, 2003) work, conceptual complexity is neither part of the definition of markedness, nor is it employed as its criterion. On the other hand, Givón (1990) does recognise cognitive complexity as a markedness criterion. The diagnostics of cognitive complexity is the degree of "*attention*, *mental effort* or *processing time*" (p. 947; original italics).[2] Thus, cognitive complexity itself is considered to be determined by extralinguistic (sub)criteria of markedness (see 2.2.6). Markedness patterns are ultimately to be explained by what Givón calls *substantive* (i.e. communicative, socio-cultural, cognitive, or neuro-biological) grounds. In all typological and functional approaches, iconicity plays an important role in connecting structural complexity to conceptual complexity,

and the cognitive notion of prominence to humans (anthropomorphic salience) is used to motivate frequency asymmetries.

2.2.3. Structural complexity

The criterion of structural complexity appears under various names in markedness research, such as *formal marking* (Lyons 1977; Winter 1989), *formal complexity* (Givón 1990), *structural coding* (Croft 1990, 2003), *markedness of symbolisation* or *morphological markedness* (Mayerthaler 1987), and similar terms. Natural Morphologists introduced the German term *merkmalhaft* to characterise a construction marked according to the criterion of structural complexity, and reserved the termed *markiert* for general markedness.

In morphosyntax, the criterion of structural complexity concerns the complexity of formal expression of category values. Croft (1990, 2003) defines this sort of complexity as the number of morphemes in word-forms in surface morphosyntactic representation: "The marked value of a grammatical category will be expressed by at least as many morphemes as is the unmarked value of that category." (Croft 2003: 92). However, other authors (e.g. Mayerthaler 1987; Haspelmath 2002) also recognise the length and complexity of a marker in terms of phonological (rather than morphological) segments as a relevant factor. A limiting case of morphosyntactic encoding is the so-called **zero** encoding, where there is no overt formal marking of a category value.

In Natural Morphology, the markedness criterion of structural complexity derives from system-independent encoding principles. Encoding of a category is evaluated for markedness in relation to these principles. The major encoding principles are: the principle of constructional iconicity, the principle of uniform encoding, and the principle of transparency. Minor principles (such as phonetic iconicity, principles of optimal word-length) are discussed in Mayerthaler (1981).

The principle of **constructional iconicity** (diagrammaticity) requires that a more marked category is encoded as more featured than a less marked category. Encoding is non-iconic if a more marked category is not encoded as more featured than a less marked category. It is counter-iconic if a more marked category is encoded as less featured than a less marked category. The principle of constructional iconicity is connected to the perceptual preference for iconic images. The principle of **uniform encoding** (biuniqueness, avoidance of allomorphy, Humboldt's universal) is based on one-to-one mapping of meaning into form. The principle of uniform encoding is connected to the

perceptual notion of object constancy. The principle of **transparency** states that a form is transparent, if it obeys the Fregean principle of (semantic) compositionality and if it is morphotactically transparent. The optimum of morphotactic transparency is the coincidence of syllable and morphological boundaries.

2.2.4. Distribution

Criteria of markedness based on distribution of linguistic values comprise behavioural criteria (termed *behavioural potential* in Croft 2003) and the criterion of neutral value. Behavioural criteria concern "any sort of evidence from the linguistic behavior of the elements in question that would demonstrate that one value of a conceptual category is grammatically more 'versatile' than the other" (Croft 2003: 95). The behavioural criteria have been developed especially in the typological approach to markedness, and their discussion here is based on Croft (1990, 2003). The criterion of neutral value concerns neutralisation of paradigmatic contrasts in certain contexts. The criterion was developed in the Prague School, and taken over by Greenberg (1966). Related is Greenberg's criterion of facultative use. Croft (1990, 2003), however, dismisses neutral value as a valid criterion of typological markedness (see below).

Two general types of behavioural potential are distinguished: inflectional (morphological) potential and distributional (syntactic) potential. The criterion of **inflectional potential** concerns the number of morphological distinctions that a particular grammatical category possesses in relation to cross-cutting (or orthogonal) categories. "If the marked value has a certain number of formal distinctions in an inflectional paradigm, then the unmarked value will have at least as many formal distinctions in the same paradigm." (Croft 2003: 97). Paradigmatic periphrastic constructions are considered to show inflectional defectivity, and hence markedness. Allomorphy and irregularity (including suppletion) are considered to be evidence for the greater inflectional potential of the category in question in the typological approach.

The criterion of **distributional potential** concerns the number of syntactic contexts in which a grammatical element can occur. Two definitions of the distributional potential are possible. A stronger definition requires that the marked element occurs in a subset of occurrences (grammatical environments, construction types) of the unmarked value. A weaker definition, which requires that the marked element occurs in a smaller number of contexts that the unmarked value, is problematic, due to the difficulty with determining

how to count distributional contexts. Context restrictions for an element may be arbitrary facts about languages, or they may stem from semantic incompatibility. Semantic restrictions are relevant for typological markedness, as typological markedness itself is a representation of constraints on the expression of conceptual categories.

The criterion of **neutral value** states that "the unmarked value is the one found in neutral contexts, where the contrast between paradigmatic alternatives does not apply" (Croft 2003: 100). According to Croft (1990, 2003), the criterion of neutral value is not a criterion for typological markedness since there is no cross-linguistic consistency as to which value is chosen, i.e. unmarked. There is no consistent cross-linguistic pattern of neutral contexts that can be linked to structural coding or behavioural potential. The explanation for this is that neutral value is, unlike structural coding or behavioural potential, not a relative concept. Nevertheless, neutral value is viewed as a valid criterion in the domain of phonology (cf. Croft 2003: 119).

2.2.5. System-dependent criteria

The proclaimed reason for introducing system-dependent criteria of naturalness (and hence markedness) in Natural Morphology is that the concept of naturalness based exclusively on system-independent factors sometimes results in incorrect predictions, especially in language change. Wurzel (1987) considers certain aspects of language-dependent normalcy to be part of naturalness. He develops two areas of system-dependent naturalness: system-congruity and inflectional class stability.

System-congruity is determined by system-defining structural properties, or typologically dominating structures in a language. System-defining structural properties of a language derive from structural properties of inflectional systems of individual word classes. The relevant parameters of system-congruity are: (a) categories (categorial systems) and category values; (b) baseform vs. stem inflection; (c) separatist vs. cumulative marking; (d) number and manner of formal distinctions in paradigms (syncretisms); (e) marker types (suffixes, prefixes, etc.); and (f) presence/absence of inflectional classes. The relatively general character of the parameters of system-congruity draws the agenda of morphological typology into determination of markedness.

Inflectional systems of a language may but need not display uniform structure. Even if there is no uniform structure in inflectional systems of a language, certain structural properties are usually dominant, and hence system-defining;

they represent set preferences for morphological structuring in the language. System-congruity then corresponds to the degree of agreement of a paradigm, an inflectional form, or a categorial marker to the respective system-defining structural properties. Properties that are less system-congruous are less natural, i.e. more marked.

The second area of system-dependent naturalness has been covered by Wurzel's (e.g. 1987, 1989, 2000) research into **inflectional class stability**. The initial observation is that inflectional classes in a language do not have an equal status: they differ in size and/or in productivity (they may gain words from other classes, or lose words). Wurzel (1989) proposes a universal principle of markedness of words regarding their inflectional class membership, which is grounded exclusively in synchronic architecture of inflectional paradigms.

Inflectional systems of individual languages are structured by so-called **paradigm structure conditions** (PSCs). PSCs are implicational structural rules (not production rules). PSCs do not necessarily operate on the basis of a uniform set of properties: their input specifications may be extra-inflectional (i.e. independently given by phonological, morphological, syntactic, and/or semantic structural properties of words in their base-forms), inflectional (i.e. category markers), or combinations of both. PSCs may be formulated for individual inflectional classes (different inflectional classes may have similar implicational structure), or generalised for the whole inflectional system of a language. The generalised PSCs determine the distribution of markers to individual paradigms (paradigm is conceived of as a specific transition through the system of PCSs), and represent canonical structures of language-specific inflectional systems.

However, not all relations between markers in paradigms and inflectional systems are implicational. Distinct inflectional classes with identical extra-inflectional properties of their members (with identical input specifications for their PSCs) are termed competing inflectional classes. One of the competing inflectional classes is usually more 'normal' in terms of type-frequency of its members. While the distribution of markers in words belonging to the more 'normal' of the competing classes follows from the (default) PSCs, markers in words belonging to the less 'normal' class must be specified in the lexicon. The existence of competing inflectional classes thus requires **lexical specification** as a further mechanism alongside the PSCs. The presence and proportion of lexical specification is considered to be the formal correlate of markedness: words whose inflection exclusively follows the PSCs (i.e. which have no lexical specification) are unmarked; words which require lexical specification of at least some of their markers are marked. An important claim is that marked-

ness in inflectional systems concerns individual markers and the words they occur in, and only indirectly also inflectional classes and paradigms (there may be inflectional classes containing both marked and unmarked words).

2.2.6. External criteria

Especially the proponents of the naturalness approach to markedness have introduced a number of extralinguistic criteria or correlates of markedness, at least programmatically. The headings in Mayerthaler (1987) include: language evolution (the later, the more marked), ontogenetic maturation (the later, the more marked), baby talk (less marked elements prefered by adults in motherese), language acquisition (less marked acquired before more marked), language disorders and speech disturbances (more marked affected/lost before less marked), perception tests (less marked more easily perceived than more marked), and error linguistics (more marked evokes more mistakes than less marked). The criterion of language change, which has been employed in several frameworks, is discussed in detail in Section 2.3.

2.3. Markedness and language change

2.3.1. The markedness reduction hypothesis

In many approaches, language change is recognised as contributing to our understanding of markedness. It has been observed that, with some qualifications, marked structures tend to develop into unmarked structures. In connection with this observation, language change has been employed as a criterion of markedness.

An absolute version of the hypothesis that diachronic developments leads to **reduction of markedness** can be easily refuted. If all diachronic developments resulted in unmarked structures, there could be no marked structures other than never-changing conservativisms. This is clearly not the case, both because all types of linguistic structures are subject to language change (although to differing degrees and retaining absolutely universal properties), and because there are innovative marked structures. Nevertheless, a number of authors have proposed some conditional version of the markedness reduction hypothesis. Assuming the hypothesis generally holds, the question is how and under what conditions marked structures arise.

In one approach, the existence of marked structures is an inevitable outcome of compartmentalisation and openness of the language system. Natural Morphologists and theoreticians of related schools of thought consider reductions in markedness to be instances of **local**, rather than global, optimisation (Dressler et al. 1987). And since different language components have diverging functions and tend to follow differing principles of naturalness, abolition of marked structures on one level leads to marked structures on another level (e.g. Lüdtke 1980; Bailey 1973). In other words, language change leads to markedness shifts from one component of the language to another, rather than to a global reduction of markedness. Markedness shifts in the above sense also exist within linguistic subsystems, due to inherent conflicts between aspects of markedness such as ease of learning and ease of perception (Bever and Langendoen 1972; Thomason and Kaufman 1988: 22–34), or various naturalness principles (Dressler et al. 1987). Marked structures are bound to exist in any moment in the history of a language. Thomason and Kaufman (1988) skeptically conclude that some (internally motivated) changes reduce markedness, some increase it, some both, and some neither.

As the markedness reduction hypothesis assumes directionality of language change, it is also connected to the issue of **teleology** in language change. Approaches allowing remedial changes have been widely criticised (e.g. Vincent 1978; Croft 2000). Abolishing the system-teleological concept of change, Stein (1989: 83) characterises directionality based on markedness in negative terms, as establishing "not teleologies, but finalities". In his view, unmarking change abolishes dispreferred structures, rather than striving for preferred structures.

2.3.2. Type of change

A different qualification of the markedness reduction hypothesis concerns the type of linguistic change involved. It is often claimed that only some types of change result in reduction of markedness or, more often, that only some types of change result in markedness increase.

One common division is between **internal** and **external** change. According to Bailey (1973), internal (or "connatural") developments, whenever language systems are left alone, is in general unmarking, while external (or "abnatural") change, resulting from contact with other systems, is said to create marked structures. Campbell's (1980) view of the connections between the type of change and markedness is unidirectional: marked structures arise only through

external change, but not all contact-induced developments increase markedness. Extending the notion of markedness from structures affected by change to mechanisms of change, Stein (1989) distinguished unmarked and marked types of change. Unmarked types of change are those that are driven by purely internal linguistic forces, while societal factors (including but not restricted to borrowing) determine marked types of change. Stein's position is, in effect, similar to Campbell's in that he considers synchronically marked states to be necessarily determined by external factors. Nichols' (1992: 249–250) exploration into macroarea-scale distribution of selected typological features reveals that, due to contact, languages with less complexity are more likely to add to it than languages are to reduce their complexity. Thomason and Kaufman (1988: 27–28), on the other hand, criticise the, in their view, simplistic correlations between type of change and markedness, arguing that both internal and external change may simplify or complicate the grammar of a language.

There have also been proposals that different **types of contact** have different consequences for the change in markedness (cf. the discussion in Thomason and Kaufman 1988: 28–32). According to Givón (1979), speakers are likely to resort to unmarked structures of the Universal Grammar in situations of contact-related communicative stress, while less stressful contact situations, those proceeding in a more gradual manner, need not involve this strategy. Bickerton (1981) predicts that his bioprogram features will resurface in contact situations of typologically diverse systems. Mühlhäusler (1980) formulates a distinction between pidgin contact situations, which lead to markedness reduction, and situations of contact between "full-fledged languages", which can result in markedness increase. According to Dressler et al. (1987), marked categories are reduced in pidgins, and unmarked categories are the first ones to be innovated during creolisation. Various creolists (e.g. Muysken 1981; Keesing 1988; DeGraff 1999) consider creole grammar to reflect the least marked parameter settings triggered by the input. Again, Thomason and Kaufman (1988) argue against simplistic correlations between type of contact and markedness. In their view, both interference through borrowing and interference through language shift may result in simplification or complication.

Nevertheless, Eckman's (1977) Markedness Differential Hypothesis suggests that contact situations involving interference through language shift should favour markedness reduction. The hypothesis states that, in second-language **acquisition**, the relative degree of difficulty for the adult learner of structures that are different in her native language and in the target language is correlated with the relative degree of markedness of these structures. Marked structures of the target language may fail to be acquired during second-

language learning, and so they may be lost in the shifting speakers' version of their new language. According to Trudgill (1989), the post-adolescent learner of a second language or a second dialect appears to be an important factor in structural simplification of high-contact varieties. Unmarked stuctures, those that are easy for the non-native speaker, are more likely to arise in large societies characterised by high contact (Croft 2000: 192–193). This seems to be compatible with Thomason and Kaufman's (1988: 32) observation that only moderate to heavy substratum interference may result in significant grammatical complications, while light interference usually produces simplification.

Thomason and Kaufman (1988: 30) refute Mühlhäusler's (1980: 28) claim that contact between dialects of a language leads to simplification of their linguistic systems. However, in a series of studies on the types of linguistic change in low-contact and high-contact varieties, Trudgill (e.g. 1989) shows convincingly that **degree of contact** and the character of social networks is a significant predictor of structural complexity, and hence markedness. Elaborating on Jakobson's observation that the wider the socio-spatial function of a dialect, the simpler its linguistic system, he shows that relatively isolated dialects and languages, which are characterised by close-knit networks among their speakers, are more likely to develop marked and redundant structures (e.g. complex phoneme segment inventories with a high number of phonological contrasts, allophonic and allomorphic complexity and morphological irregularity, complex agreement patterns). On the other hand, high-contact dialect contexts (such as dialect mixture and koinéisation), which are characterised by relatively open social networks, are likely to produce unmarked linguistic structures through decrease of irregularity, redundancy, and complexity.

2.3.3. Markedness and language contact

Although **borrowing** as such has been claimed to increase markedness (see Section 2.3.2), markedness is also differentially operative in borrowing. More specifically, markedness co-determines what forms and functions are more likely to be borrowed than others. The discussion in a recent state-of-the-art study on language contact, Winford (2003), distinguishes various types of constraints on borrowing, among them constraints based on markedness/naturalness and transparency (itself one of the naturalness principles in Natural Morphology). In the view of Thomason and Kaufman's (1988) claim that there are no absolute constraints on borrowing, it may be more appropriate to speak of factors rather than constraints.

The markedness factors Winford (2003: 92, 94–96) summarises from various sources (e.g. Weinreich 1953; Heath 1978; Thomason and Kaufman 1988; Dalton-Puffer 1996) are mostly **structural** properties of borrowed elements in the context of the source language, especially properties concerning form–function mapping. They include the degree of boundedness and overall integration, explicitness of marking, the degree of variation in form, unifunctionality and categorial clarity, as well as other aspects of form–function transparency of the loan element. The more transparent and the less integrated an element is in the source language, the more likely it is to be borrowed.

Similar structural factors have been adduced in order to explain observed preferences in pidgin/creole formation. The selection of superstrate forms is believed to be determined by perceptual salience and, again, transparency (Siegel 1999), both of which may be subsumed under markedness (Winford 2003: 345). Identical structural factors presumably play a role in other contact situations that involve imperfect learning (cf. Bardovi-Harlig 1987 for the role of salience in second-language acquisition). Mufwene (1991) argues that markedness is relative to the context of the contact situation of pidgin/creole formation: an unmarked typological option in pidgin/creole formation need not correspond to the unmarked option in the lexifier language. He believes that, in determining the degree of markedness of a marker, salience is a criterion superior to semantic transparency. Analytic markers are said to be more salient than synthetic markers as they are more easily emphasised; and, salient synthetic marking is more likely to be retained than less salient synthetic marking. Frequency and occurrence in most of the varieties in contact are yet other factors that define unmarked strategies in pidgin/creole formation.

Mufwene (1990) also suggests that markedness in pidgin/creole formation is relative to typological distance between systems in contact. Typological distance, especially congruence in morphological structures, is another type of factor Winford (2003) postulates for borrowing, too.

The third and final type of constraints on borrowing in Winford (2003: 96–97), **functional** factors, are considered to be of minor importance. For Winford, functional factors in borrowing concern categorial additions and losses, including filling in functional gaps, and are distinct from markedness factors. However, markedness clearly plays an important role in determining what functions and categories are more sensitive to borrowing. In his paper on local markedness, for example, Tiersma (1982) notes that while generally it is the singular form of a noun loan that is adopted, with nouns referring to natural groups or pairs it is the plural form. Categorial markedness has also been employed in studies on second-language acquisition (Berretta 1995).

Stein (1989) extends the notion of markedness not only to mechanisms of change (see 2.3.2) but also to modes of diffusion throughcontexts, i.e. to **actualisation** patterns. He claims that while in internal change innovations enter in the least marked points and extend to more marked environments, in borrowing and other externally motivated change innovations spread from marked points to less marked contexts.

Chapter 3
Toward a communication-based model of asymmetry in language

3.1. Factors involved in the formation of asymmetry

Markedness, we noted above, presupposes that the criteria employed to describe values do not act independently. Rather, the assumption is that different criteria will tend to cluster. Furthermore, it is assumed that the assignment of properties to one pole on the continuum, rather than the other, is not accidental, but conditioned by perceptions of reality. We also noted various hypotheses regarding markedness and language change. A case is sometimes made for a correlation between 'marked' values and participation in changes that lead for instance to greater complexity, while 'unmarked' values show greater susceptibility to changes that lead, for instance, to greater differentation, simplification or erosion. Our sample offers an opportunity to revisit the markedness reduction hypothesis by examining changes that have taken place within the time depth postulated since the breakup of Early Romani. In particular, the Romani sample gives us an excellent opportunity to examine the role of so-called external factors, notably language contact, in change.

We share the view that the phenomenon known as 'markedness' is essentially a phenomenon of asymmetry among members of a paradigm. But for the moment we would like to put aside categorical assignments of the labels 'marked' and 'unmarked' to individual values within a paradigm. This applies both to oppositions in the binary sense, but also to poles on the markedness continuum. Instead, we would like to examine markedness as a *local* instance of asymmetry involving values of a particular category paradigm, and particular processes that affect the shape of these values. We assume that these processes are motivated by global factors that shape communication. We view the so-called 'perception of reality', which is said to be reflected in language, through the functional prism of communication: The structure of communication mirrors aspects of the structure of reality, as perceived by the participants in communication, since making use of familiar categorisations is an effective way to guide the hearer through the processing of propositional contents and illocutions presented by the speaker. As Croft (2003: 116) puts it: "Assuming

that human beings must master the structure of experience, it is more efficient that language parallel that structure as much as possible". While we accept that asymmetries are not accidental, but shaped by cognitive universals, we believe that the usefulness of asymmetry is not to portray reality as such, but to facilitate communication about reality. Consequently, our goal is to investigate the communicative functions of asymmetry.

The communicative motivations that give rise to asymmetry are the key to understanding why there is no permanent, pre-determined assignment of 'markedness' or 'unmarkedness' to any given value in a paradigm. Asymmetries are rather a product of the interplay of several factors in a specific structural environment; and a structural 'environment' is a conventionalised device for achieving certain communicative objectives. We will enumerate those factors here, then revisit them from the perspective of their role in communication.

The first factor is what we call **Category**. This is the common denominator of **values** in a paradigm, and a reflection of real-world categorisations in the structuring of information. The categories we shall explore include Person, Negation, Number, Cardinality, Modality, Gender, Animacy, Tense, Case, and more. The values of a category are the options available for making a statement about the status of a structure in relation to the specific category. Thus, the category Person has the values first, second and third person (participant roles, and third entities); the category Cardinality differentiates between higher and lower ordinal and cardinal numbers; and the category Degree has the values positive, comparative, and superlative. The overt representation and arrangement of values is language-specific, but the values themselves reflect conceptual categorisations of reality.

Categories and values are represented in language at the level of linguistic **structures**, which are language-internal procedures for organising information at the level of the utterances and the discourse. Pronouns, for instance, are structures that organise information about continuous referents. This information is typically categorised in Romani according to Person, Number, Gender, Animacy, Case, and Discreteness. Verbs are structures that organise information on events and actions, and are typically categorised in Romani according to Transitivity (valency or participant structure), Tense, Aspect, Modality, and, in relation to the main participants in the event or action, also according to Person, Number, and occasionally also Gender.

The key to identifying asymmetry is to compare the way different values of the same category are expressed within individual structures. In carrying out such comparison, we apply **criteria for asymmetry**. These include processes that give rise to the formal representation of values, such as Erosion, Extension,

or Borrowing, as well as the outcomes of processes, such as Complexity, Differentiation, or cross-dialectal Diversity (representing general susceptibility to processes of renewal); we shall discuss them in more detail toward the end of this chapter. Some models of markedness take the view that relevant criteria will tend to cluster, so that the application of any criterion to a value will render the same picture of markedness. We refer to this approach as a 'static' approach to markedness. In the 'static' approach, criteria that do not fit within the cluster are rejected as irrelevant to the overall identification of markedness, even if they too lead to asymmetries. We take, by contrast, a 'dynamic' approach to markedness. We do not view markedness as a pre-determined presence or absence of a cluster of relevant properties for each value. Rather, we expect variation in the patterns of asymmetry, since we view them as contextually determined, arising from an interplay of several factors. The processes that are behind the criteria for markedness reflect strategies, by which speakers draw connections between the properties that are attached to a particular value, and the communicative objectives that are being pursued through the employment of a particular structure. These strategies are employed in a way that is advantageous to the needs of communication, and so they tend to draw on elements of the cognitive-conceptual categorisation of states-of-affairs in the real world.

We therefore assume that there are communicative advantages to creating and maintaining asymmetrical relationships between values of individual categories in particular structures. These advantages trigger what we call the **conceptual motivations** for the formation and maintenance of asymmetry. Conceptual motivations are grounded in conversational maxims that ensure communicative efficiency by assigning hierarchical status to chunks of information, thereby assisting the hearer to prioritise the processing of information in discourse. Since the motivations for maintaining aymmetry are conversational, we identify the relevant dimensions in terms that relate to the conversational status of the information that is being prioritised via an asymmetric structure, e.g. Transparency and Accessibility, Relevance, Saliency, and Egocentricity.

In our 'dynamic', communication-based model of markedness, then, concrete asymmetries are determined by an interplay of (a) structures, which convey information, (b) the category in relation to which this information is being contextualised, and the relationships between individual values of this category, (c) a range of conceptual motivations to prioritise information, drawing on the grid of relations between these values, and (d) various strategies by which priorities can be mapped onto formal structures. Let us review the role of each of these factors, with reference to the model depicted in Figure 3.1. The pri-

mary objective of asymmetry is to aid in the effective management of communication. Communication management is partly achieved already through the packaging of information into structures. These are the **structures** to which we have referred above, and which constitute formal structural categories of language (nouns, pronouns, verbs, clause combining structures, and so on). Structures contribute to the effective organisation of communication by specialising in particular types of information-depiction, e.g. verbs depict events and actions, pronouns portray participants, complex constructions portray relations between events, and so on. Although divisions into word-classes and types of clause are language-specific, they show many universal properties, and are ultimately anchored in cognitive universals of perception of real-world events.

One way of increasing the efficiency of information conveyed by structures is to combine the structural division with another dimension, by contextualising information through meaningful categorisations. These are what we referred to above as **categories**. For example, actions may be linked to their

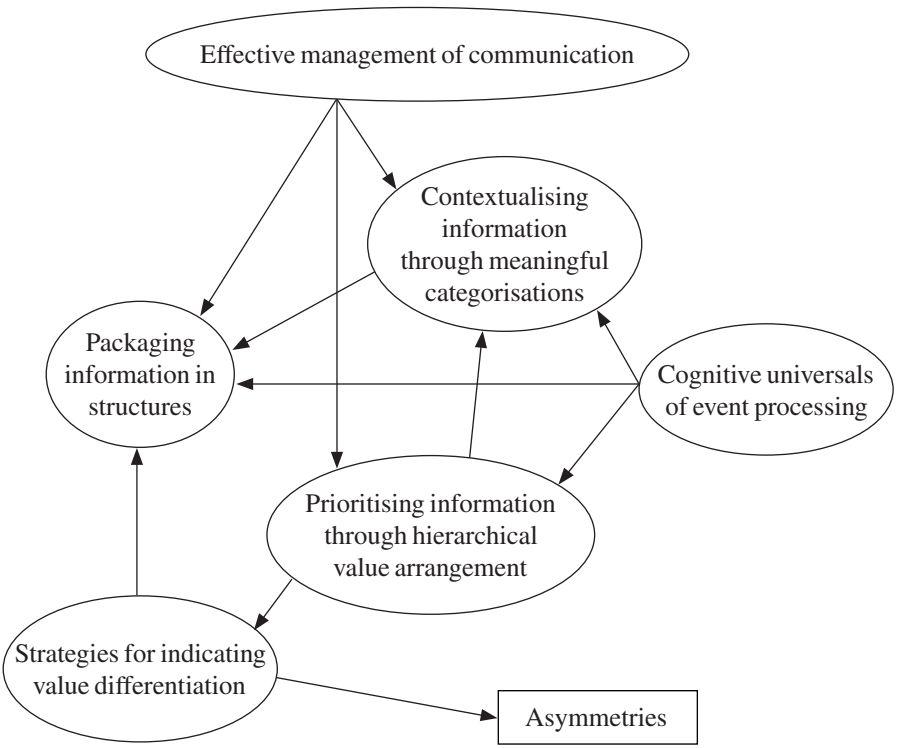

Effective management of communication
Contextualising information through meaningful categorisations
Packaging information in structures
Cognitive universals of event processing
Prioritising information through hierarchical value arrangement
Strategies for indicating value differentiation
Asymmetries

Figure 3.1. The communication-based model of asymmetry in category paradigms

participants (by combining the structure 'verb' with the categories 'person' or 'number' or 'gender'), the relations between events may be linked to expectations of prototypical connections between events (by marking out the degree of discreteness, or contrast, between clauses). Meaningful categorisations are of course language-specific, but they too are inspired by cognitive universals of event processing.

Now, the efficiency of communication management may be increased yet further by prioritising portions of information through a hierarchical arrangement of values within meaningful categories. The reference grid for such prioritisation of values remains the framework of meaningful categorisations, or categories. The borderlines that are set between individual values of these categories are, once again, language-specific, but they too are inspired by cognitive universals of event processing and real-life experience, pertaining to the conceptual evaluation of the specific values. This is why we have referred to the advantages of prioritising information in this way as **conceptual motivations** to discriminate between values.

The motivations to prioritise values inspire the design of concrete strategies of value differentiation. We have mentioned above that in the analysis we draw on such strategies and their outcomes as **criteria** for assessing asymmetry among values of a given category. Strategies of value differentiation are of course bound to a specific structure, since it is via structures that communicative goals are achieved. It is through the effect of these strategies that asymmetries become apparent and detectable in the structures of language.

On a communication-oriented model of language, then, asymmetries are *strategies for differentiating structures, in order to prioritise individual values of categorised information, in a way that is based on the experiences of processing real-world events*.

3.2. Application of the model

Consider three examples for the application the above model:

Example (1). In managing communication, it is advantageous to alert the hearer to relations between events, and to monitor and direct the hearer's reaction to the links which the speaker creates between positions in the discourse. This is achieved, at the level of the interaction, through discourse marking devices, and at the level of linking propositions, through **structures** of clause combining. These structures are contextualised by linking them to categorisations of relations between real-world events. We refer here to the **category**

of Discreteness, which we understand as the evaluation of two entities as belonging together, or as constituting separate and unrelated entities. Values associated with Discreteness are, among other things, addition (independent entities are conjoined together), and contrast (independent entities are conjoined together contrary to expectation). Now, in order to prioritise information processing, contrast is singled out as a pole in the set of values of the category of Discreteness.

The **conceptual motivation** for targeting contrast is related to the considerable mental effort that is needed on the part of the speaker in order to obtain the cooperation of the hearer when asking the hearer to accept an unexpected link between events, and to the fact that, more than in other points on the Discreteness continuum, the speaker's authority is at stake. Specifically, contrast is a high priority because of its high degree of relevance to existing assumptions shared by the speaker and the hearer.

The concrete strategy of indicating a hierarchical relationship between contrast and other values of the category Discreteness, which we often observe in Romani varieties, is the borrowing of a contrastive conjunction. Whereas lower-ranking values may be borrowed, the conjunction 'but' is always borrowed. Based on the **criterion** of borrowing, we can therefore detect an asymmetry in the category of Discreteness, as expressed within the structure of clause combining, between the value 'contrast', and other values.

In this particular case, it is the specific factor of bilingualism that directs the speaker to the strategy of borrowing. Following the argument presented in Matras (1998b), we assume that one way for the bilingual speaker to prioritise contrast is to eliminate the burden of having to select the appropriate linguistic expression from among the two sub-components (= 'languages') within their linguistic repertoire. This reduces somewhat the tension surrounding the planning of the utterance in those instances in which tension arises due to the clash between anticipated hearer-expectations, and the nature of the link between propositions which the speaker is about to present. Contrast, in other words, creates tension in the linguistic-mental planning procedure; bilingualism is responsible for a permanent tension wherever the speaker must follow rules on the appropriateness of choices among sets within his/her overall linguistic repertoire. In order to ease tension around contrast, the choice among sets is eliminated through the process of 'fusion' (see Matras 1998b), by which a single set of expressions is generalised for use in both languages. Since Romani is a group-internal language, and the Romani expression cannot be used outside the Romani language due the sociolinguistic contraints imposed by the majority language, it is the minority language, where bilingualism is accepted,

that gives way, adopting the set of the majority language. This is the conversational-structural background for borrowing in the case of contrastive conjunctions and, as we shall see, in the case of numerous expressions and morphosyntactic devices that are high on the relevance scale.

Example (2). In communication, it is advantageous to avoid ambiguity of reference to salient participants. Pronouns are a structure that allows the speaker to keep track of topical entities. This is done by linking pronouns to the category Person, a procedure by which participants in the discourse, namely speaker and hearer, can be identified and kept distinct from non-participants (third persons). The individual persons are thus values of the category Person, which, inspired by action-related roles in the real world, is encoded in the structure of pronouns. Now, in increasing the efficiency of disambiguation, there is a need to prioritise third persons. This is because to those who are present in the interaction, the identities of speaker and hearer are apparent already through the reference to their discourse roles. With third-person entities, some potential for ambiguity remains, triggering a conceptual motivation to reduce ambiguity and maximise transparency. In Romani, various strategies are involved, among them greater differentiation of third-person pronouns (which unlike the other persons inflect for gender, and whose nominative forms are arguably irregular), and through more frequent renewal of the expressions used to refer to third persons, often by recruiting deictic expressions, leading to greater diversity among the expressions used for third-person entities among the dialects.

Example (3). As we shall see in more detail in Chapter 5, Early Romani had inherited a differentiated set of reflexes of the Middle Indo-Aryan participle marker *-t-*, which indicates the formation of the perfective stem in verbs. Following the voiced dental sonorants *r, l, n* as well as *v* it shows voice assimilation, giving *-d-*: *ker-d-o* ‘done’. Following vowels, the dental stop shifts to a dental lateral, giving *-l-*: *xa-l-o* ‘eaten’. Elsewhere, we can assume continuation of **-t-*: **dikh-t-o* ‘seen’. The outcome was an early differentiation into three distinct morphological classes of perfective markers – in *-d-*, *-l-*, and *-t-*. In the later Early Romani period, however, a tendency appears to have emerged to avoid certain consonant clusters resulting from the attachment of the old perfective marker *-t-* to consonantal verb stems. The solution to the articulatory tension that the clusters create is to re-assign the relevant verb stems to a different morphological class, namely to the class in *-l-*, which originally had included only vocalic stems. This class re-assignment, a morphological solution to an articulatory problem, is found in most of the dialects, but to different extents. Dialect comparison allows us to trace the hierarchical progression.

The cases which demand earlier solutions, and so which are more widespread across the dialects, are those where the clash resulting from dissimilar articulations was most extreme: the combinations **mt, *gt*, **kt*, more so than **čt*, or the even more permissible **št*. On a phonetic hierarchy of obstruents vs. fricatives, the historical participial marker in an obstruent **-t-* tends to be avoided in positions next to other obstruents: *-t* > *-m* > *-g*, *-k*, *-kh* > *-č*, *-čh* > *-š*, *-s*.

Stems in *-t* are rare in the language. Those that can be found belong exclusively to the perfective inflection class in *-l-*: *xut-l-* for *xut-* 'to jump'. Only the most conservative dialects still show traces of the *-t-* marker with stems in *-m-*: Welsh Romani *kam-d-om* 'I wanted', with late voicing, Latvian Romani *kam-dž-om* alongside *kam-j-om* < **kam-lj-om*. By contrast, forms in *-t-* have the highest survival rate in positions following sibilants. The hierarchy is implicational, and we find that if a dialect has preserved *-t-* in stems ending in *k*, for instance, then it will also preserve it in stems ending in sounds that are lower on the hierarchy (cf. Polish Romani *mukh-t'-om* 'I left', *phuč-t'-om* 'I asked', *beš-t'-om* 'I sat'; Epiros Romani *mukh-lj-om* 'I left', *phuč-lj-om* 'I asked', but *beš-tj-om* 'I sat').

This phoneme hierarchy is difficult to reconcile with any typical hierarchy of phonological markedness. The extended perfective marker, *-l-*, is of course higher on the sonority hierarchy than either of the alternative morphemes, *-t-* or *-d-*, so that it has an articulatory advantage in the environment of obstruents. Phonological hierarchies relate to a different set of motivations than morphosyntactic phenomena, since the former do not involve conceptualisations (cf. Croft 2003); we will not be dealing with phonology in the data chapters of this book. But this example underlines how strategies are employed (the extension of a morpheme, in this case) at the local level, in order to make the procedure of information packaging more efficient (in this case, in order to facilitate articulation).

3.3. Criteria for asymmetry

We have mentioned that we regard asymmetries in linguistic structures as the outcome of strategies that are employed in order to prioritise information in communication. The grid for prioritisation is the paradigm of values within a category; and categories in turn help contextualise the information presented by linguistic structures (word classes and types of phrases and clauses). Based on this communicative-functional understanding of asymmetry, our interest is not, primarily, in determining which value is 'marked' and which

is 'unmarked', but in exploring the interplay of values of categories, structures, and the strategies themselves. Identifying the strategies is the key toward gaining an understanding of the underlying motivations to prioritise individual values, in individual structures; and this in turn may allow us an insight into how categorisations of real-world events and states-of-affairs help shape language. We can identify strategies, since they are directly involved in shaping the formal structures that package information. This gives us a catalogue of **criteria** for identifying asymmetry among the values of a category in a given structure. In this section, we survey the criteria which we apply in our investigation.

3.3.1. Complexity

Complexity is normally regarded as the number of morphemes in the morphosyntactic surface representation of word-forms. Thus, words or constructions that are composed of two morphemes are more complex than ones that are monomorphemic (e.g. English *dog-s* > *dog*, *more common* > *common*) etc. Some authors also recognise the length and complexity of a morpheme in terms of its phonological segments as a relevant factor (see Chapter 2). If asymmetries in phonological complexity are salient and regular, we take them into consideration. For example, we consider Romani demonstrative inflections, which generally consist of two phonemes, a consonant and a vowel (e.g. *-va*, *-ja*, *-la*), to be more complex than adjective inflections, which consist of a single phoneme, a vowel (e.g. *-o*, *-i*, *-e*). Marginally, we also assume some phonemes to be 'stronger' or more complex than others (e.g. *s* > *h*, in Chapters 7 and 13).

Functionally speaking, complexity can be interpreted as a strategy to single out individual values, relying on the combined effect of more than one morpheme in order to narrow down the structural representation; cf. English *this one*, as opposed to *this one over here*. Alternatively, it may be regarded as an iconic representation of real-world complexity or real-world inaccessibility; cf. the addition of the Romani suffix *-as* to denote remoteness, as in *džav* 'I go', *džav-as* 'I was going/ used to go'.

3.3.2. Erosion

Erosion is the diachronic reduction of phonological segments or morphemes, the shortening of forms, or fusion of boundaries. In English, there is an erosion hierarchy involving the reduction of *it is* to *it's* in the present tense, but

not usually in the past tense (*'*twas*, except in literary contexts). Erosion is essentially a process of simplification, and the opposite process from the one that leads to complexity. Nonetheless, the two are not opposite poles of the same phenomenon. While erosion always leads to loss of complexity, the rise of complexity has no effect on erosion. Erosion is therefore linked unilaterally to complexity, while complexity is not linked to erosion.

The most straightforward asymmetrical effect of erosion occurs when markers of a category under investigation are reduced to differing degrees. For example, in the category of case (see Chapter 16), the Early Romani genitive marker undergoes radical erosion *-*ker*- > -*kr*- > -*k*- > -*č*- > 0 in some dialects, while the other overt oblique case markers lose a single phoneme at most (e.g. dative *-*ke* > -*če* > -*e*, ablative *-*tar* > -*ta*). Erosion, however, may also induce asymmetries in a category without affecting markers of that category. For example in interrogatives, word-initial erosion is more likely to affect the manner interrogative **sar* > (*h*)*ar* 'how' than the interrogative determiner **savo* > (*h*)*avo* 'which, what sort of' (see Chapter 20 for details). Here, the segment affected by erosion is the marker of the interrogative function, the interrogative root **s*-, not the markers of the 'ontological' distinction between manner (i.e. -*ar*) and determiner (i.e. -*av-o*). Although erosion has an asymmetrical effect in the 'ontological' category, it does not affect markers of that category. Finally, asymmetrical erosion may, of course, also affect phonological segments that belong to more than one morpheme. For example, the sequence of the 1SG non-perfective marker -*av*- and a following tense marker is more likely to contract in the imperfect than in the present-future: cf. Welsh Romani contracted imperfect *ker-ās* (< **ker-av-as*) 'I used to do/ was doing', but uncontracted present-future *ker-av-a* 'I (will) do' (see Chapter 13 for details).

3.3.3. Differentiation

Differentiation is identified in relation to distinctions of cross-cutting categories, by taking into account the number of cross-cutting categories that are relevant for a value, and the number of values within each of those cross-cutting categories. For example, within the category Person, we find, in the structure of Romani pronouns, that the value of the third person (i.e. the third-person pronoun) is differentiated for the cross-cutting categories of gender, number, and case. Within the same structure of pronouns, a different value of the category Person, namely the first person, is only differentiated for number and case, but not for gender (compare this with the similar state of affairs in English *he*/*she*/

it/*they* and *him*/*her*/*it*/*them*, but *I*/*we* and *me*/*us*). The value third person is thus more differentiated. Differentiation is measured by the inflectional potential of the value. A differentiated value is one that potentially inflects for more cross-cutting categories, with more values of those cross-cutting categories being explicitly distinguished. Under differentiation we also examine the depth of differentiation, or the irregularity of the relationship between individual values of cross-cutting categories. Greater irregularity between the values means greater differentiation (cf. English *he*/*him*, where the cross-cutting value case in the value third person is comparatively regular, drawing on the same consonantal root of the pronoun, whereas in *I*/*me* the cross-cutting value case in the value first person is irregular, drawing on distinct consonantal roots of the pronoun for the subject and object cases). Irregularity as a criterion for markedness is thus subsumed in our analysis under Differentiation. Functionally, differentiation may be regarded as a strategy that allows narrowing down the pool of potential referents or events, by encoding more detailed information about them, thereby contributing to disambiguation.

3.3.4. Extension

Extension is a diachronic process by which the marker of a value (either the entire form, or some influence from a form) becomes the marker of another value within the same category. There are, in principle, two distinct types of extension. In the first, the marker of one value is extended to denote also a second value, and as a result the two values become formally indistiguishable (cf. English possessive *her* extending to object). In the second type, a formal distinction is maintained, either since the extension was partial (with one form influencing but not completely replacing another), or because, following a full extension and replacement of form B by form A, a new structure C emerged, taking over the function previously covered by A, so that instead of an oppostion A/B, we now have an opposition C/A. We thus identify extension wherever there is movement of form from one value within a category, to another value within the same category. Analytically speaking, movement from one value (A) to another (B) allows us to reconstruct a hierarchical relationship between the values, such that if A has been extended, then it was a value that was given priority, in some way, at the expense of B.

Functionally, extension may serve several purposes. First, it indicates the prominence of a value by turning it into the point of orientation within its paradigm, with other values accommodating to it. Second, it gives priority to

a value that is considered broader in its potential scope of meaning, so that it can be interpreted as covering a wider range of values. This value is allowed to 'move on' in the paradigm, and to express the meaning and function of neighbouring values, at the expense of weaker values. A pre-requisite for the extension of the 'stronger' ('unmarked') value here is its ability to take on contiguous meanings in a non-literal sense, at least initially, and so not just its referential but also its metonymical flexibility. By this definition, many cases of grammaticalisation and semantic bleaching qualify as cases for extension (cf. colloquial British English *wicked* meaning 'exceptional' or 'noteworthy').

3.3.5. Extra-categorial distribution

Distribution is the syntactic side of the so-called behavioural potential of a value (Croft 2003: 95), and we follow this and other sources in looking for the versatility of a form in terms of the syntactic contexts in which it appears. Unlike extension, which we have defined as a movement from one value within a category to another value, extracategorial distribution implies movement from one category to another category. This covers, for instance, the frequent use of interrogative forms in other functions, such as conjunctions (*kaj* as a relativiser and factual complementiser). Admittedly, the boundaries of categories are not always entirely clear-cut and objective, and their definition is sometimes subject to purely analytical considerations. For example, we choose to separate the set of negative indefinites (such as 'nobody', 'nowhere', etc.) from other expressions of negation, and to treat them instead as a value within the category of indefiniteness. This is because there is no obvious polarity between negative and 'positive' indefinites (thus, 'nowhere' is not the negative opposite of either 'somewhere', 'anywhere', or 'everywhere'), so that the more relevant consideration in evaluating negative indefinites is their position within a set of indefinites, rather than their opposition to positive counterparts. Perhaps a somewhat more arbitrary analytical categorisation, one that we take primarily for reasons of convenience, is the distinction between localisation (with which we mean the expression of spatial relations, and its extension to temporal dimensions), and case roles (which are the roles assigned to actors in the utterance, via prepositions and inflectional case). Bearing this in mind, recognising extra-categorial distribution relies on the demarcation of categories.

An example of extra-categorial distribution in Romani morphosyntax is the use of masculine singular inflection markers with items that do not inflect for gender and number, such as indefinite pronouns or interrogatives (cf. *khonik*

'somebody, anybody', OBL *khanik-as*, or *kon* 'who', OBL *kas* 'whom', both irrespective of the gender or number of the potentially intended referent), indicating that the masculine singular is used outside its categories of indicating a specific value in the category sets Gender resp. Number. A further example is the use of the 3SG subject concord marker on the present-subjunctive verb (*-el*) as a so-called 'new infinitive' in some dialects to denote agreement at the level of the complex clause, between the subject of the modal complement and that of the main clause or matrix verb, irrespective of the actual identity of that subject (e.g. the subject of both clauses may be a first or second person, or plural). Employed in this way, the 3SG is no longer a value in the categories Person resp. Number; rather, it encodes a non-finite verb form, and has thus been extended into a sub-group of the category of 'converbs' or specific devices marking subject continuity across clauses.

Functionally, extra-categorial distribution provides a standardised device for contextualising information, so conventionalised that it is also employed under the heading of another category. The choice of a particular standardised device (or value) reflects the prioritising of a particular value of the original category for this purpose. This offers an opportunity to draw on more familiar values for more specialised uses, thereby generalising from one context to another, which in turn allows economising the use of specialised expressions or forms by relying on the context for any differentiated processing.

3.3.6. Exposition

By exposition we refer to the unique and consistent formal representation of a value. There are three possible formats for exposition. The first is a material format: A value is more exposed because it is more consistently distinguishable, for instance by a larger number of distinctive features. Thus, the 1SG concord marker in *ker-av* 'I do' is more exposed that either the 2SG *ker-es* 'you do' or the 3SG *ker-el* 'he/she does', since it has both a unique consonant (*-v*) and a unique vowel (*-a-*), whereas the other two forms have unique consonants, *-s* and *-l* respectively, but share a vowel (*-e-*). In an alternative format for exposition, the value of a form is always predictable, because the value is neither extended, nor extended upon. For example, in the form *kerdjam* 'we did', *-am* is always recognisable, both within individual dialects, and, as it happens in this case also cross-dialectally, as the 1PL. Other values of the paradigm may be extended or partially extended (e.g. the extension of the vowel *-e-* and absence of jotation, from the 3PL *kerd-e* 'they did' to the 2PL *kerdj-an*

→ *kerd-en*), or extended upon (cf. the position of 2PL in the same example).

In a third possible format for exposition, the form of a value is always predictable, because the value is never extended upon; but the value of the form is not always predictable, since it may extend. For example, in cross-dialectal comparison of the Romani definite article forms, the masculine singular (= value) is always *o* (= form), but *o* is not always restricted to the masculine singular, since it may extend to other values (as in Velingrad Yerli, where there is a uniform definite article *o* for all genders/numbers), and because *o* may also result from erosion of the plural definite article *ol*.

Because of this latter chance factor, the only consistent definition of exposition is the more restrictive one, namely that by which a value is exposed if it is not extended upon (i.e. not subsumed by another value). Exposition will thus mean that the value is formally set apart to a maximum degree from any other values in the paradigm. In the category Degree, for instance, the positive is always exposed, having a separate form from both the comparative and superlative (e.g. *baro* 'big'), while the two latter values may sometimes be expressed by the same form (e.g. *bareder* 'bigger, biggest'). Returning to the masculine singular form of the definite article *o*, it is the most exposed because, in any given dialect, it is most likely to be different from all other forms of the definite article (it happens also to be the least diverse, so that predictability of the form is very high; but this cross-dialectal predictability is not in itself the criterion for exposition).

Unlike extension, where one form has several values, exposition means a consistent form–function (form–value) relationship. But absence of extension is just one possible factor leading to exposition; another, as in the 1SG *ker-av*, is the distinctiveness that is asserted through a maximum number of features. Functionally, then, what exposition achieves is a kind of referential stability which sets apart the expression of a value from that of any other values.

3.3.7. Borrowing

As noted in the previous chapter, borrowing is often considered to be an extralinguistic factor involved in markedness. Although we acknowledge that, from a system point of view, borrowing obviously involves the incorporation and replication of system-external elements, we regard borrowing primarily as the outcome of particular circumstances of communication management, in which speakers negotiate boundaries between sets of structures within their overall linguistic repertoire. In numerous contexts and sets of contexts, speakers will

be guided by the rules of appropriateness, well-formedness or simply just for the sake of communicative efficiency to restrict their choices to just one set of structures – one 'language', as it were, employed in a monolingual context. This is of course the case for speakers of Romani when using their second language(s) in a context determined by the majority language and culture.

The situation is more complex in group-internal communication. The choice within the group is for Romani, i.e., for the set of structures that are identified as exclusive to the Romani context, and so as appropriate, and indeed in some ways even constitutive, of the Romani interaction context. However, there are communicative advantages of various kinds to the lifting of boundaries or constraints on the employment of selected elements from the non-Romani set. This may involve the extension of vocabulary to denote artefacts, actions and institutions that are typically situated outside of the Romani community. It may also pertain to entire utterances or parts of utterances, where code-switching contributes to the organisation of discourse (see already Gumperz 1982; Auer 1984).

Our focus however is on grammatical borrowing, and we regard these as instances where speakers permanently lift the mental 'tag' that identifies a form as belonging to a particular (in this case, the non-Romani) set, and so create a licence to employ this form on a regular basis, irrespective of the characterisation of the context and the constraints created by the context on choices within the repertoire. Borrowings are therefore items that can be etymologised as belonging to another system originally, but which are in permanent and regular use within the Romani system, and no longer subject to context-bound choices between sub-sets of the repertoire. In many cases, grammatical borrowings can be assumed, or even proven, to be 'replacive' borrowings, which have substituted functionally equivalent expressions inherited from earlier stages of the language. Above we noted some advantages of certain types of grammatical borrowing, precisely in eliminating the need to choose between sub-sets of the repertoire (or 'languages') in expressions that already trigger high processing tension due to their position on the relevance scale, and the special challenge of maintaining communicative harmony with the hearer. There are, quite possibly, other advantages to borrowing, as well, and examples will be discussed in the second part of the book.

Borrowing of form or concrete matter is sometimes distinguished from the replication of just patterns, also referred to as loan-shifts (Weinreich 1953), metatypy (Ross 1996), or convergence (Silva-Corvalán 1994; Matras 1998b). We will use 'borrowing' to refer strictly to the 'transfer', so to speak, of concrete morphemes or morphological 'matter'. This is because convergence, or

the transfer of patterns, is present in almost every domain of morphosyntax in Romani, and is often involved in triggering such changes as differentiation, extra-categorial distribution, or extension, with which we deal separately as independent criteria.

3.3.8. Internal diversity

The advantage of a cross-dialectal sample is the ability to make generalisation on the basis of recurring patterns, and not just on the basis of the structural composition and distribution of values of a category in one variety. While this is a strong argument in favour of samples that are cross-linguistic, the unique advantage of a cross-dialectal sample is that it allows us to conclude, at least on numerous if not on all occasions, that diversity within the sample is a product of language change, and not just of a coincidental distribution of different options of expressing the same functions or meanings. Diversity in itself, then, can be a useful criterion with which one can assess proneness to renewal and change. Quite often, diversity will be an outcome of proneness to borrowing, with dialects each adopting forms from their respective contact languages. We therefore tend to discuss diversity in conjunction with borrowing in the individual chapters. However, we try to pay special attention to internal diversity, caused by a range of processes including internal grammaticalisation, extension, erosion, reduplication, and others. Diversity is thus a by-product of other criteria for asymmetry, and, unlike the other criteria that are employed, not one that directly reflects strategies of structural formation and categorisation through hierarchisation, but one that allows us to identify susceptibility to such strategies indirectly.

3.3.9. Criteria not included in this study

One of the central criteria for markedness employed in the literature is frequency (see already Greenberg 1966). Croft (2003: 110ff.) discusses the difficulties in employing frequency as a criterion, noting in particular the distinction between frequency of the denoted entity in the real world, and text frequency, which is the frequency with which human beings choose to communicate about these entities in conversation. Greenberg had argued for a causal relation between frequency and structural simplicity, based on the assumption of a default use of unmarked forms. We are unable to include frequency, due to

limitations on our corpus (see Chapter 4). But irrespective of those limitations, it is also questionable whether frequency has any direct manifestations that are observable in structural change, and so useful in a cross-dialectal comparison. It is not clear, for instance, whether frequency should be the trigger of change (we speak about a value more frequently than before, hence the value becomes, in Greenberg's terms, 'unmarked', and so it tends to adopt a more simple structure, undergoing some kind of erosion, perhaps), or whether it is merely an indicator of structural change (simplified structures are easier to use, and so their frequency increases).

Another criterion which we do not employ is that of conceptual complexity. As we explained above, we regard conceptual motivations as the trigger behind structural strategies of prioritising information. In a sense, the conceptual complexity of a value, relative to other values of the category, is the reason why a value might be singled out structurally, or 'marked'. Conceptual considerations themselves can therefore not be taken as indicators of markedness, since that would make the argument circular.

Due to limitations imposed by the data corpus, we are also unable to take into consideration so-called system-external criteria such as speech production errors, child language acquisition, aphasia or other language disorders, or, in the case of Romani, the rather hypothetical, or at least extremely rare constellation of second-language acquisition.

3.4. Factors motivating asymmetry

We said above that there are conceptual motivations to prioritise certain categorial values that are associated with pieces of information. We view these motivations as connected intrinsically to the speaker's attempt to win over the hearer's solidarity and agreement in communication. Conceptual factors that motivate the structure of communication are derived from the cognitive organisation of information on the real-world, but their purpose in communication is to draw on these categorisations in order to establish a shared point of view with the listener. In this respect, the incorporation of cognitive categorisations into discourse is functionally motivated: the speaker is motivated to prioritise information in order to construct linguistic interaction in an efficient and effective way, and the speaker is also motivated to draw on existing cognitive categorisations when prioritising information, in order to match the priorities in discourse to those of the real world. In this section we review some of the factors that are involved in motivating the setting of priorities in discourse.

3.4.1. Topical saliency

When transferring information in communication we pay special attention to entities that are in the centre of natural events, the initiators or undergoers of events, typically those with whom we can identify, or whose actions or states have some bearing on us. The structure of language may reflect strategies to prioritise information on those entities. Topicality is generally understood as the degree of prominence that is given to referential entities (cf. Givón 1984, 1990). It is linked to attention, and to the status of other pieces of information as modifying, explaining or in some other way enhancing our understanding of topics. Topicality, then, is the 'aboutness' of an utterance. The saliency of topics is, in turn, the degree to which potential topical entities are given attention compared to other potential topics. Topical saliency is thus a hierarchical status in a given discourse environment. However, due to the nature of discourse as focusing to a large extent on relevant human activities, topical saliency is to some extent predictable in terms of a combination of various semantic scales. Thus, prominence is likely to be given to humans over non-humans, to animates over non-animates, to agentive and intentional actors over non-agentive ones, and to known entities over unknown ones.

3.4.2. Transparency

We regard transparency at two levels: First, it is the stability or consistency of a form-meaning relationship as expressed by a linguistic structure. More transparent structures are those that are more autonomous, in that their meaning is less dependent on changing contextual environments. Lexical items are in this respect more transparent than grammatical modifiers, and derivational morphology is more transparent than inflectional morphology. Second, transparency is the ability to decode information in an unambiguous way. Liability to change, as in the expression of events or actions, may be less transparent than static properties, as expressed by nouns. Categorisations of objects by means of a static relation to just one additional object (for example through local relations *next to*, *behind*) may be more transparent than their categorisation in relation to more than one object (*between*). Depiction of temporal relations (categorisation of an event in relation to a specified point in time) is likely to be more transparent than one involving processes of cause and result, where a more complex set of potentially changing circumstances must be taken into account.

3.4.3. Discourse accessibility

Closely related to transparency is the notion of discourse accessibliity. Here too, we are dealing with the ability to avoid potential ambiguity and so to ensure harmony between speaker and hearer in identifying events and referents, and establishing points of view. Whereas transparency should be viewed as an inherent semantic property of an independent structure, we understand accessibility as dependent to a greater degree on discourse factors, and so as a property that emerges when structures are evaluated in context. The accessibility of referents is dependent on the hearer's ability to identify entities pointed out by the speaker. Thus, the participants in an interaction, speaker and hearer, are more easily accessible than third entities that do not have an active role in the interaction. Nearer objects and ongoing events are more easily identifiable than remote objects or hypothetical or conditional events. Continuity is an important factor determining accessibility, as continuous entities are more easiliy retrievable than discontinuous ones.

3.4.4. Egocentricity

Egocentricity is a reflection of the self-centred character of cognition. Its origin may be seen in the instinctive pre-occupation with the self, with language as an expression of this pre-occupation. We understand the role of egocentricity in conversation as the attempt to integrate the hearer into the speaker's own and immediate point of view. Deixis and the role of the speaker's self in determining the deictic centre is one reflection of egocentricity, as is the ranking of the hearer as second to the speaker, followed by third entities. Egocentricity mixes with accessibility in the priority that is given to events and objects that are nearer and more immediate, over those that are remote and concluded.

3.4.5. Relevance

Relevance structures are discourse strategies and grammatical devices that serve functions of strengthening or contradicting existing assumptions, and so confirming the relevance of assumptions (Sperber and Wilson 1986). They require an assessment of discourse presuppositions, and so an intensive monitoring of hearer-sided knowledge and processing. Relevance thus relates to information structure, at all levels, especially at the discourse level. Structural

devices that are high on the relevance scale are those that involve more explicit processing of hearer-sided assumptions, and especially the contradiction of hearer-sided assumptions, for this is where the speaker faces a special challenge to win over the sympathy and solidarity of the hearer.

3.5. Concluding remarks

We have described as 'static' those models of markedness that attempt to make general predictions about the overall markedness of individual values of categories. Such an approach to markedness will inevitably lead to the exclusion of some criteria for markedness which do not fit in with the general pattern, but which may nevertheless indicate other asymmetries that are relevant to the discussion of asymmetry in the respective category. As an alternative, we have proposed what we referred to as a dynamic, communication-oriented model of asymmetry. This model views asymmetry as the outcome of local solutions to local challenges triggered by the interplay of structural packaging of information, contextualisation of this information through categorisation, the need to prioritise information provided in this categorised framework, and conceptual motivations to prioritise this information in a fashion that is compatible with priorities set by real-life experience. We predict that these motivations play a central role in triggering and in regulating processes of change in language.

In the second part of the book, we review our cross-dialectal data sample by category, applying to the categories criteria for asymmetry based on the outcomes of common processes of change. We take into account the (emergence of) complexity and differentiation, extension, erosion, extra-categorial distribution, and exposition, In addition, we pay attention to borrowing, which is such an outstanding feature of Romani, as well as to internal diversity, an indicator that is available to us through our reliance on a cross-dialectal sample.

Chapter 4
The sample: Methodological considerations

4.1. Sampling in a typological context

The objective of constructing a typological sample is to enable linguists to make generalisations about human language without considering all individual languages, i.e. through inference. The reason for constructing a typological sample is, of course, the various limitations a typologist is bound to encounter, if s/he wanted to investigate all languages. The limitations include (from the more practical to the more theoretical ones): finance, time, capacity, lack of description, and the inaccessibility of most extinct and all future/potential languages.

Typologists may wish to arrive at at least three sorts of **generalisations**: (a) generalisations about the range of linguistic variation; (b) generalisations about the relative frequency of linguistic phenomena; and (c) generalisations about preferences and dispreferences in human language. Different types of samples may be constructed accordingly (see below). The latter two sorts of generalisations are by no means identical (Dryer 1989). A major development in typological sampling has been to develop techniques to arrive at generalisations about preferences in the human linguistic potential rather than about frequencies contingent on the current linguistic situation.

There are two major theoretical requirements for a sound typological sample: the sample should be representative and, at the same time, independence of instances should be ensured. The conflicting nature of these requirements has triggered important theoretical discussions on the subject; at the same time, it also appears to create unsolvable problems (Croft 1995; Song 2001) (see below).

Representability of a typological sample is required in order to minimalise inferential errors in formulating typological generalisations. A sound sample should neither under-represent nor over-represent languages with regard to (some of) their extralinguistic properties. Most researchers have focused on identifying and avoiding genetic and areal biases, although other sources of biased samples have been identified too: bibliographical accessibility (Bell 1978); age of the language (the failure to include pidgins and creoles); medium (the failure to include sign languages, Comrie 1993: 8); or contactness (the failure to include highly contactive languages, Elšík 2001). Apart from

these extralinguistic factors, some authors (Comrie 1981; Stassen 1985; Dryer 1989) have suggested controlling for major (highly predictive) typological parameters of sample languages, too.

In terms of representability, a universal sample seems to be the ideal sample. Nevertheless, this assumes that the sum of all at least potentially accessible (i.e. extant and attested extinct) languages is indeed representative of the human linguistic potential. This assumption has been questioned. It cannot be excluded that all current languages descend from a common ancestor ("Proto-World", Comrie 1981), or that they participate in a global linguistic area (Dryer 1989). Moreover, the current languages may be simply lucky, being the only ones to have survived due to social and/or technological dominance of their speakers.

The requirement of **independence of instances** has been pursued especially from a statistical perspective (Perkins 1989). First, the selection of sample units should be statistically independent of irrelevant variables. Perkins assumes that, for example, the number of speakers of a sample language, or the number of languages in a genetic or areal grouping to which a sample language belongs, are irrelevant for structural typology. According to this assumption, a universal sample is not a good sample since it is bound to overrepresent features of large genetic groupings or large areas.

Second, the relevant properties of sample units should be statistically independent of one another. The relevant properties include structural features of a sample language as well as, for example, its genetic and areal affiliation. This requirement complicates any attempts at avoiding typological bias. According to Perkins, it has to be tested whether an a priori stratification of sample languages according to a major typological parameter does not hinder the statistical independence of the phenomenon under investigation. Perkins has applied statistical tests of "association" (i.e. absence of independence) to some previously constructed large samples (Maddieson 1984; Tomlin 1986; Ruhlen 1987), and found out that they would need to be reduced significantly to fulfil the requirement of statistical independence.

Thus, it appears that the two major requirements on typological samples are in an inherent **conflict**. On the one hand, representability requires a maximal number of languages to be included in the sample; such a sample, however, is bound to contain statistically dependent instances. On the other hand, independence of instances requires constructing samples with a small number of languages; such samples, however, are bound not to be representative.

Some authors (Dryer 1989; Croft 1990: 22–23; Comrie 1993) have been more optimistic with regard to the conflict of the principal requirements, noting

that a lack of statistical independence does not automatically imply a lack of **historical** independence. That is, linguistic features of genetically and/or areally related languages may be historically independent. Indeed, all differences among closely related languages have been claimed to be historically independent for principled reasons (Comrie 1993). Thus, provided historical independence may be shown, even statistically dependent instances may testify to universals of the human linguistic potential.

Finally, we consider different types of (stratified) typological **samples**. According to a sample's structure one may distinguish: (a) convenience samples, which may try to avoid major biases but necessarily remain largely arbitrary; (b) proportional samples (e.g. Bell 1978; Tomlin 1986), which aim at proportional representation of genetic or other groupings and so provide generalisations about frequencies of linguistic phenomena, but cannot say anything about linguistic preferences; and (c) hierarchical samples. The latter may be exemplified by Dryer's (1989) sample, which operates with the hierarchy of macroareas and 'genera' (i.e. genetic units roughly corresponding in time depth to individual branches of Indo-European).

According to the goal of a sample one may distinguish: (a) variety samples (e.g. Rijkhoff et al. 1993), which aim at discovering the possible variation on a linguistic parameter, and whose emphasis is thus on representability; and (b) probability samples (e.g. Dryer 1989), which aim at formulating statistical universals and typological correlations, and whose emphasis is thus on independence of instances. The procedure of **double** sampling appears to be an attractive option: first, one constructs a larger variety sample to learn about the typological range of the phenomenon under investigation, and about its distribution among language groupings of various sorts; and second, a smaller, statistically tested, probability sample is constructed in order to formulate significant typological generalisations.

4.2. Dialect sampling in Romani

4.2.1. The usefulness of dialect samples

Dialect sampling has traditionally been concerned with detecting patterns of geographical and social diffusion of innovations. More recently, however, a new, typologically oriented dialectological research context has emerged (see Kortmann 2004). Its agenda is to sample dialects for syntactic-typological features, in order to pool descriptive data on syntax and investigate syntactic

phenomena not found in standard languages, to allow a fuller description of dialect structure than those provided by traditional dialect studies, and to investigate developments in syntax that are not constrained by written language. In pursuit of these objectives, we have seen in recent years the emergence of several databases devoted to dialect structure.

Beyond just syntax, typology can support the study of dialectal variation even more generally, by helping us to examine how universal functions of language are categorised in any specific variety. The two fields of study can be combined most effectively when typological methodology is applied to comparative dialectology. A sample of related dialects can help us gain insights into mechanisms of language change in a group of related varieties. We are interested here in the degree of diversity vs. uniformity, and the question of which structures are susceptible to change. But we are also able to capture change during its diffusion across dialects, in order to examine how an emerging innovation is distributed across values of a category in its intermediate stages.

Consider for example the erosion of the Early Romani middle marker -(*j*)*ov*- and its reduction, ultimately giving rise to just a vowel component which marks an additional inflection class. Table 4.1 shows the inflections for middle verbs in some dialects (see Chapter 7 for more details).

The general picture is clearly that of a person hierarchy: -*jov*- is most susceptible to reduction in the third person, followed by the second person, which in turn also shows a number split whereby the plural is more susceptible to reduction that the singular, followed then by the first person (see Section 7.2 for a more detailed hierarchy). But this hierarchy is only evident by comparing the varieties. The dialect sample thus allows us not just to discover variations on a linguistic parameter, but it allows us to determine the probable relations among values within a paradigm.

Table 4.1. Middle inflections in selected dialects

	3SG/PL	2PL	2SG	1SG/PL
Rumelian R	*-ov-e-/-o-*	*-ov-e-/-o-*	*-ov-e-/-o-*	*-ov-a-*
Piedmontese Sinti	*-o-*	*-ov-e-*	*-ov-e-*	*-ov-a-*
West Slovak R	*-o-*	*-o-*	*-ov-e-*	*-ov-a-*
Roman	*-o-*	*-o-*	*-oj-*	*-oj-a-*
Varna Kalajdži	*-o-*	*-o-*	*-o-*	*-ov-a-*
Ajia Varvara	*-o-*	*-o-*	*-o-*	*-a-*

One of the specific features of Romani dialects is the fact that their common point of departure can be placed at a time depth of no more then seven centuries, and that subsequent change has not been constrained by any institutional norm. In the next section, we outline the special challenge of Romani.

4.2.2. The challenge of Romani

Romani is the only Indo-Aryan language that is spoken exclusively in Europe.[1] It derives from an idiom formed in the transition period between Middle and New Indo-Aryan, which we might call 'Proto-Romani' (see Matras 2002). This idiom, and its contemporary descendants, present-day Romani dialects, show some very conservative traits in phonology, some of them even pre-dating Middle Indo-Aryan.[2] There are also archaic features in nominal and verbal morphology, such as the retention of consonantal case suffixes on the noun, and of the Middle Indo-Aryan present-tense conjugation of verbs. Other features are characteristic of the drift toward New Indo-Aryan: the reduction of the nominal case system to a two-way opposition nominative vs. oblique, the emergence of enclitic case markers, and traces of an ergative system that followed the generalisation of past participles and later collapsed.

Proto-Romani was carried from India westwards by migrants who appear to have been members of service-providing castes, similar in status and occupational profile to the jatis or profession groups known in some parts of India as *ḍom*. This term, *ḍom*, is even cognate with the self-appellation *řom* used by the Rom (or Romani) of Europe, as well as with the term *dom* used by so-called Gypsies in the Middle East who also speak an Indian language.[3] The *řom* settled in the Byzantine Empire, sometime around the tenth century CE, where their language absorbed a strong Greek element in lexicon, syntax, phonology and even inflectional morphology. 'Early Romani', as one might term the stage of the language prior to the collapse of the Byzantine Empire and the emigration of some groups via the Balkans into western and northern parts of Europe, can be regarded as a *Balkanised Indo-Aryan language* (see already Matras 1994b).

Romani migrations from the Balkans are recorded from the fourteenth century onwards. Romani populations began to settle throughout Europe around the late fifteenth or early sixteenth century, often maintaining itinerant occupations, albeit within more or less fixed territorial boundaries. The earliest samples of Romani are from the mid-sixteenth, the seventeenth and early eighteenth centuries. On the whole these already represent the kind of dialectal

variation that is known to us from the present-day dialects of the language. The isoglosses that represent internal (i.e. not contact-induced) innovations within the shared or inherited component of Romani tend to form geographical patterns (see Matras 2002, Ch. 9), and we can therefore assume that (a) many, if not most of the features separating the dialects emerged in situ, spreading into geographically neighbouring communities, and (b) the bulk of these developments took place between the early period of settlement, from the fifteenth century onwards, and the period from which wider documentation is already available (and dialectal variation as manifest in present-day dialects is already encountered), i.e. the mid-eighteenth century. While individual isoglosses form many different geographical patterns, a rather dense cluster of isoglosses can be found separating south-eastern Europe from the Northwest. This line has Transylvania, Vojvodina, and Croatia as its northern frontier, corresponding roughly to the political boundaries which existed at least during a significant part of the relevant period between the Ottoman and Austrian (Habsburg) empires.

The unique structural and sociolinguistic position of Romani makes it a fascinating case-study, especially in respect of the interface of typology and dialectology: Firstly, there is no form of standard Romani, even though written publications and internet correspondence in individual dialects of the language have been expanding rapidly since the early 1990s. There is not even a globally accepted prestige dialect. Every form of Romani is therefore a 'dialect'. Not only is there no generally accepted standard, there is also rarely any normative influence on the individual dialects.[4] Romani has usually been an oral language, and this is still the case for the great majority of the Romani-speaking population. Even the emergence of oral media in the language is a very recent phenomenon, and one that is strongly contained and restricted to the local level, in just a few localities. Bias in favour of a standard, addressed by Kortmann (1999) as a potential problem of typological methodology, can therefore not occur when dealing with Romani.

Next, since Romani has always been an oral language, and since written historical records of it are lacking (with the exception of the rather late samples referred to above), historical reconstruction of the language relies entirely on comparative dialectology. When comparing dialects, one often has to draw on universal features of grammaticalisation processes in order to be able to identify which structures are older, and which are innovations.

For example, the origin of the definite article in Romani has been subject to some debate. Some authors of older descriptions had speculated that the Romani preposed article, which is inflected for gender and number, and in

some dialects shows the forms M.SG *o*, F.SG *i*, PL *e*, is a Greek borrowing. Romani is unique among the Indo-Aryan languages in having preposed definite articles, and their emergence in the language will have certainly been triggered by contact with Greek, although direct borrowing of the forms can be ruled out. Some dialects possess forms in *l-*, mainly for the oblique and/or the plural, and some have forms in *ol-*. This suggests an affinity to the paradigm of third-person pronouns, M.SG *ov*, F.SG *oj*, PL *on*/*ol*, which also has oblique forms in *l-* (and more rarely in *ol-*). In fact, the remote demonstratives also contain a component in *-ov-*, *-oj-* and *-ol-* for the same categories, with additional preposed and sometimes also postposed segments.

Typology makes us aware of a universal cycle of reduction of the deictic force and of movement across the categories along the path *demonstratives* → *third-person pronouns* (*anaphora*) → *definite articles* (see Himmelmann 1997; Diessel 1999; and already Greenberg 1978). This helps us reconstruct the original forms of the definite article as identical to the third-person pronouns *ov, oj, ol*, and oblique *ol-*, which in turn emerged from identical predecessor forms of the contemporary remote demonstrative (*od-ov-a, od-oj-a, od-ol-a* and so on) (see Table 4.2).

Third, there is a special challenge in applying dialectological methodology to a language that lacks a geographically coherent continuum of speaker populations, and where migrations continue to play a significant role in settlement patterns. The fact that non-territorial languages may show geographical diffusion of innovations in a way that is similar to territorial languages has been shown for case of Yiddish by the Weinreich dialectological project (Herzog, Weinreich, and Baviskar 1992). For Romani, geographical plotting of isoglosses is a method applied only recently, mainly at the regional level (cf. Boretzky 1999a, 1999b). A diffusion model is arguably applicable to Romani as a whole, and can be in fact quite useful for the purposes of dialect classification (see Matras 2002, Ch. 9).

Table 4.2. Reconstructed Early Romani determiners

	M.SG	F.SG	PL	OBL
Original remote demonstratives	*ova*	*oja*	*ola*	*ol-*
Emphatic demonstratives	*od-ova*	*od-oja*	*od-ola*	*od-ol-*
	ok-ova	*ok-oja*	*ok-ola*	*ok-ol-*
Anaphoric pronouns	*ov*	*oj*	*ol* (*on*)	(*o*)*l-*
Definite articles	*o*	*e* (*i*)	*o*, *ol*, *le*	(*o*)*l-*

Finally, Romani shows a number of exceptional sociolinguistic features: Romani serves in the first instance as a symbol of identity and a vehicle of group-internal or even family-internal communication. All adult speakers of Romani are bilingual. Due to the socio-economic structure of the Romani community as, primarily, a service economy, economic transactions are generally negotiated outside the Romani community, in the majority language. This is reflected in the structure of the language: It has a relatively small core vocabulary that is independent of borrowings from current or recent contact languages, and it is on the whole extremely absorbent not only of borrowed vocabulary but also of borrowed grammar, including derivational and even inflectional morphology. Since speakers are always bilingual, the sample of Romani dialects displays the structural outcomes of contact with a great variety of languages, under very similar sociolinguistic conditions of contact, with no exposure to normative influence. Moreover, many dialects have, in the course of their history, absorbed successive layers of influences from successive contact languages. Romani therefore offers a chance to study general and perhaps even universal mechanisms of contact-induced change, under almost ideal comparative conditions.

4.2.3. Romani dialectology

The bulk of the research work carried out on Romani is descriptive in nature, concentrating on outlines of the phonology, morphology and lexicon of individual dialects. More recent descriptions tend to devote chapters also to syntax, as well as to contact influences (see e.g. Holzinger 1993; Boretzky 1994 Igla 1996; Halwachs 1998; Cech and Heinschink 1999). There are also several studies that examine individual phenomena in Romani in cross-dialectal as well as typological perspective (e.g. Boretzky 1993b on conditional sentences, Matras 1998a on deictics, Elšík 2001a on indefinites, Koptjevskaja-Tamm 2000 on genitives, Matras 2001 on tense–aspect–modality, Crevels and Bakker 2000 on external possession; see also Matras 2002 for a comparative discussion of Romani dialects).

Dialect classification in Romani has its roots in Miklosich's (1872–1880) comparative survey and attempt at a historical discussion. Miklosich's classification was based on a reconstruction of the migration routes of those groups that left the southern Balkans. The principal reference features were not innovations with the internal, inherited component, but rather the successive layers of loan vocabulary. The result was a reconstruction of the branching and

sub-branching of groups from several main waves of migration, a grid that later inspired Romani dialectologists to postulate several dialect branches, and so a 'genetic' split. A different kind of approach to dialect classification in Romani was taken by Gilliat-Smith (1915), focusing on the dialects of north-eastern Bulgaria. While Miklosich emphasised the in situ character of the dialects, Gilliat-Smith recognised that in the area under his investigation, dialects belonging to different branches overlap geographically. The distinction between 'settled' and 'nomadic' dialects had already been introduced for the Balkans by Paspati (1870). Gilliat-Smith adopted the term *vlax*, used by immigrant (mainly Orthodox, and nomadic) Rom originating from Wallachia, contrasting it on a wholesale basis with the *non-vlax* (mainly Muslim, and settled) Romani populations. Due to the frequent presence of immigrant communities speaking Vlax Romani dialects in other parts of Europe, for a while this distinction was adopted as a kind of 'basic' dialect division within Romani. Occasionally, authors still characterise a particular dialect as being 'non-Vlax', even if it is spoken in a remote location from the Vlax dialects, has never been argued to be a Vlax dialect, is not otherwise a candidate for a Vlax dialect, and so does not really need to be argued, explicitly, to be a non-Vlax dialect.

Gradually, however, a division into several dialect groups of equal ranking emerged, which became a popular reference grid in work on Romani linguistics during the 1990s (cf. Bakker and Matras 1997; Elšík 2000c). This division recognises a Vlax branch (centred around the historical Wallachian and Transylvanian regions, with outmigrants in various regions throughout Europe), a Central branch (with a northern sub-division in southern Poland, northern Slovakia, and Transcarpathian Ukraine, and a southern sub-division in southern Slovakia, Hungary, eastern Austria and northern Slovenia), and a Balkan branch (including the Black Sea coast dialects, and occasionally sub-divided into a 'default' Balkan dialect – Southern Balkan I in Boretzky's (1999a) terminology – and a distinct sub-group based in northeastern Bulgaria and Macedonia – called Southern Balkan II or the Bugurdži–Kalajdži–Drindari group). More controversial are the status and affiliation of the dialects of western and northern Europe, including southern Italy and the Iberian peninsula. Bakker (1999), following other suggestions in the literature, had grouped them together under the heading of a so-called 'Northern' branch. It seems more realistic to define separate Northeastern (Baltic) and Northwestern (German–Scandinavian) groups, and to treat the remaining dialects as isolates (see Matras 2002, Ch. 9).

In the centre of the controversy surrounding any classification model is the question of whether a feature that is shared by several dialects can be regard-

ed as 'genetic'. This is the approach taken at least by some studies of the late 1990s, which adopt a pre-defined group of dialects, and then work their way inwards, enumerating the features that are common to the group, then extending the comparison to individual features shared with other groups (see Boretzky 1999a, 1999b, Bakker 1999). The predominant notion is that a density of shared features represents a historically coherent population group that spread over a larger territory as a result of migration, while a limited inventory of features shared with another group represents earlier ties with that group, prior to the break-away through migration of one of the populations. At the same time, however, Boretzky (1999a, 1999b) introduced into Romani dialectology the plotting of feature inventories on regional maps – an admission that geographical patterning of isoglosses could indicate that diversity emerged in situ, through the geographical diffusion of innovations, rather than just through the dislocation of populations and consequent 'genetic' subbranching.

The approach adopted in Matras (2002) places the emphasis on the diffusion of innovations through larger geographical spaces, and the patterns of larger-space isoglosses that emerge. In interpreting these patterns, emphasis is given to central methodological notions in dialectology such as the distinction between archaisms and innovations, and between centre and periphery. The resulting picture allows us to identify two primary centres of innovation in Romani. The first is in the north-west, and its centre is in or around Germany. The second is in the south-east, though two distinct types might be recognised: The first covers the entire Balkans; the second is more specific to Transylvania/Wallachia, but often influences the Balkans, especially the Black Sea coast, thus sub-dividing the southern Balkans into an eastern and a western zone. The two major centres are separated by the Great Divide – a bundle of isoglosses alluded to above, and following roughly the line Croatia–Vojvodina–Transylvania–Wallachia. They contain features such as the assimilation of the third-person singular of past-tense intransitive verbs into the transitive paradigm in the north, the prothesis of *j-* in third-person pronouns and lexical words in the north, the consistent analogous renewal of the oblique form of the interrogative 'who', *kas*, to *kon-es* in the north, the retention of *ov-* as an auxiliary in the south, the use of an (originally participial) extension *-in-* with the copula stem in the south, the loss of the nasal in the nominal suffix *-pen/-ben* in the south, and several more. The primary classification of an individual dialect can thus be in relation to its participation in a particular isogloss development, which would mean in relation to its historical geographical location, or more precisely to its location during the period of isogloss formation.

The eastern division of the Balkans, following roughly a line from Transylvania which then separates the Black Sea coast region in Bulgaria from the west of that country, shows features such as the retention of a cluster in historical *ndř*, prothesis of *a-* in many lexemes, affrication of palatalised dentals as in *cikno* 'small' and *dzes* 'day', renewal of demonstratives through addition of a suffix in *-k-*, retention of loan verb adaptation markers in *-is-* or *-iz-*, and more.

The reference grid or so-called 'consensus classification' has a partial reality, however, in that it tends to represent the clustering of a number of isoglosses that have to do with the re-structuring of a number of complex morphological paradigms, notably demonstratives, loan verb adaptation markers, and analogies in the set of past-tense concord markers. In relation to this particular cluster of isoglosses, one might tentatively speak of a consistent grid of features allowing separation of the Balkan, Vlax, Central, Northeastern, and Northwestern dialects.

4.3. Putting typology to work in a dialect sample: The Romani Morphosyntactic Database (RMS)

In this section we outline the Romani Morphosyntactic (RMS) database. The database has information on around 90 different dialects of Romani. From its onset in late 1999, the project had two aims. First, to record the extent of variation in Romani, focusing on morphosyntax, using an electronic format to facilitate access to data (through search facilities, data import and export functions, and so on). Second, to use the sample of Romani data to evaluate the extent and the nature of diversity (and so historical change) in individual morphosyntactic functions of language. In plain terms, the first question is dialectological. Its focus is primarily formal-structural: The concrete representation of an underlying or historical structure in variety X. The second question is typological, its focus being on the structural representation in variety X of a universal function of language.

4.3.1. The database tools

The RMS database has a user interface that displays fields in distinct layouts, each covering a grammatical chapter (such as 'Indefinites', 'Articles', 'Case Representation', 'Word Order', 'Adverbial Subordination', and so on). The

browser may thus use the database exactly like a reference grammar, which sub-divides phenomena, evaluates them by providing answers to analytical questions (e.g. on the etymology of items, on morpho-phonological modifications, on extensions of historical forms to cover other functions or categories), and of course exemplifies them by displaying concrete data.

The database has essentially two types of entries that are accommodated in the data fields. The first are data entries, which exemplify phenomena directly from the variety under investigation. Here, the user enters the shape of words, grammatical affixes, or sentences as recorded in the dialect. The second type of entry is the analytical entry. The purpose of the analytical entry is to evaluate and classify data, and so to categorise the dialect with respect to any specific phenomenon. Analytical entries are typically assigned value lists, which anticipate relevant answers and allow the user entering data to select the answer that is relevant to the variety under investigation, or, to add the appropriate value. Values may be answers to analytical questions: typically either yes/no questions (e.g. 'Is the definite article retained?' → 'y/n'), or questions relating to the function or distribution of a structural category (e.g. 'What is the function of short present forms?' → 'subjunctive, present/subjunctive'; 'Does the lexical verb occur in final position in adverbial clauses?' → 'attested/unattested/facultative/obligatory'; 'What is the case of the possessor ?' → 'NOM/ACC/DAT' etc.). Value lists often contain concrete structures which are classified as types. Thus for instance, listing the historical forms rather than the concrete representation of the form in the individual dialect, the type of preposition might be asked in connection with the way a particular local relation is expressed. This allows the user to identify the semantic-functional scope of an inherited preposition, and so to compare the dialects in relation to the way they express a function, as well as to trace the semantic-functional development of an inherited form across the dialects (whereas the concrete phonological shape of that preposition in an individual dialect is listed elsewhere).

Value lists also identify the etymology of forms that are prone to borrowing, by allowing a choice between 'inherited' and several layers of 'L2-borrowings': The 'Old L2' represents a contact language that has played a role in the history of the dialect, but is no longer spoken as a second language in the community (an example being Rumanian influence on the Vlax dialects of Serbia). The 'Recent L2' represents a contact language that is no longer spoken by the younger generation of a Romani community, but may still be known to the older generation (an example is Hungarian for speakers of Roman, the Romani dialect of the Austrian Burgenland district). 'Current L2' represents the principal contact language, spoken by all generations in a given community. A key

to the identity of the individual contact layers is provided in a special 'Profile' layout, for each of the dialect sources. Individual dialects can, of course, have more than one contact language in a single layer. Thus for Arli in Macedonia, both Macedonian and Albanian might be regarded where applicable as 'Current L2s', and so on. The tagging of etymology for depth of contact allows a stratification of contact influences, which in turn may allow insights into the susceptibility of individual categories and function to replacive borrowing or generally to contact-related change.

4.3.2. Function to form, form to function

Above, two questions were described: The dialectological, examining the distribution and concrete representation of the language-specific structures of Romani, and the typological, examining the structural representation of specific functions of language in individual varieties of Romani. Superficially one might equate the first question with a descriptive procedure leading from *form* to *form* (for instance the concrete phonological shape in variety X of the inherited Romani dative marker **-ke/-ge*), or in some cases from *form* to *function* (for example the functional scope of the inherited syncopated present-tense form, possible options being the Present Subjunctive, or the Present Indicative+Subjunctive, as opposed to the long form of the present in *-a*, options being Present/Future, Future, Conditional, and so on). The second, typological question might be equated with a procedure from *function* to *form*. For instance, the continuum of semantic integration of complement clauses is captured by a range of main clause predicates representing tighter and less tight event integration (such as *can*, *want*, *begin*, *try*, and *fear*), as well as the contrast between modality (*can*, *begin*, etc.) and epistemic complementation (*see*, *know*, *hear* etc.), and between identical subject and different-subject constructions (so-called manipulative predicates such as *demand*, and *ask*).

In practice, the two strategies – form to function, and function to form – are often integrated. Thus, if we take complementation as an example, we find several dimensions. First, the aforementioned range of predicates, representing points on a continuum of semantic integration and semantic control. This follows typological work on complementation, as represented for instance by the works of Wierzbicka (1988), Givón (1990), Frajzyngier (1991), Frajzyngier and Jasperson (1991), and Dixon (1995). For each predicate, three value lists appear. The first contains a statement about the presence or absence of a complementiser conjoining the two clauses. The value options are 'none', or a

choice of a complementiser type. This latter value is a Romani-specific form. Modal complements tend to take a non-factual complementiser of the type *TE* (realised in the individual dialects as *te*, *tə* or *ti*). Epistemic complements tend to take a complementiser of the type *KAJ*, though this latter is often replaced by a borrowed particle. (The pre-defined value lists operate on the basis of expected variants. However, any value list can also be amended by the inputter to include a value that has not been anticipated.)

The next field identifies the origin of the complementiser, the value options being 'non-applicable' (in case a complementiser is absent), 'inherited', or a choice between several layers of borrowing (see above). The following field characterises the inflection of the complement verb. The value options are 'finite' and 'non-finite'. Clause combining in Romani is overwhelmingly finite. However, in modal complements with identical subject constructions ('infinitive clauses'), some (mainly central European) dialects tend to generalise one of the person-inflected forms, thereby abandoning subject agreement, and introducing instead a kind of 'infinitive', based historically on one of the finite forms. The final field is a data field, into which an example is inserted.

Figure 4.1 shows an example of entries for the *Yerli* dialect as spoken in Velingrad, Bulgaria (acquired for the database through direct elicitation). With

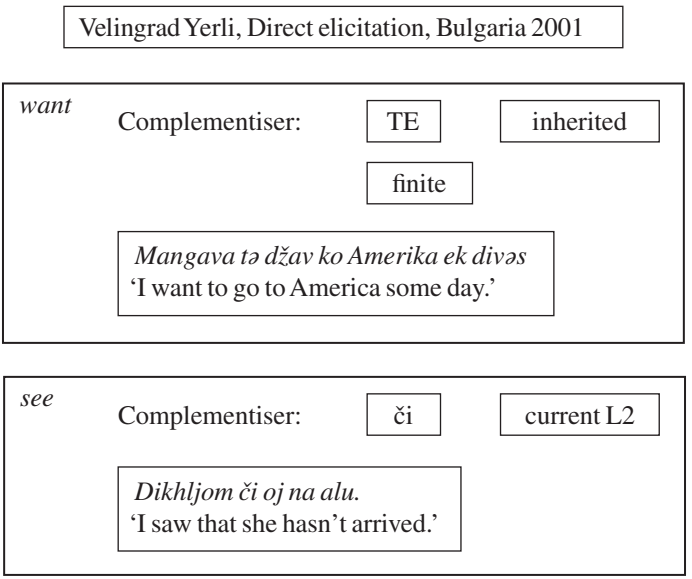


Figure 4.1. Sample entries for complementation

the modal verb *want*, the complementiser is *tə*, historically **te*, and so *TE* is the type selected from the value list. The etymology field indicates that it is inherited (and so part of the pre-European component). The complement verb is finite, showing person agreement with the subject of the matrix clause, and the absence of the present/future suffix *-a* marks it out for the subjunctive: *dža-v* 'go-1SG'; cf. the matrix verb *mang-av-a* 'want-1SG-PRES'. For the verb *see* we find a different state of affairs. The complementiser *či* is borrowed, and so the concrete form is entered. The etymology field indicates a borrowing from the current contact language, which for this dialect is Bulgarian. The question of the finiteness of the verb is redundant in epistemic constructions, where no Romani dialect uses non-finite forms, and therefore it does not appear in the entry.

The evaluation possibilities offered by this form of organisation are both functional, and formal-structural. The user can examine the categorisation of functions of language in the form of concrete structural patterns in the variety under scrutiny. In the case of complementation, the relevant categorisation involves different means through which clause integration is expressed: the employment of connectors, or the choice of a particular pattern of verb inflection. Equally from a universal-functional viewpoint, the distribution of inherited as opposed to borrowed complementisers, and the stability of borrowed forms, can be assessed by categorising predicates according to the etymology of the clause combining particle which they take. This in turn enables the user to observe possible tendencies for bilingualism and language contact to trigger change in particular functions. From the internal Romani viewpoint, dialects can be grouped according to the means of clause combining which they employ in particular functions (here: combinations of specific predicates with a complement). This may involve the specific form of a complementiser, or the presence or absence of one, or the presence or absence of subject agreement (= finite verb) on the complement, and so on. The Velingrad Yerli example in Figure 4.1 shows first of all a split in the choice of a complementiser between the two matrix predicates, and the proneness to borrowing of the second complementiser, that introducing the complement to the epistemic predicate *see*.

The form-to-function perspective is, in comparison, very straightforward. As mentioned, Romani dialects inherit two forms of the present stem: A short form, in which the final morpheme indicate person concord (1SG *-av* etc.), and a long form, where the suffix *-a* attaches to the person concord morpheme. It appears that the long form served as a present-future in Early Romani, while the short form was the subjunctive. The dialects continue both forms, but alter

Table 4.3. Inherited present-stem forms and their TAM function in some dialects

Dialect	Short form	Long form	Analytic future
Sepečides	subjunctive	present	*ka*
Rumelian R	subjunctive	present	*ka*(*m*)
Kosovo Bugurdži	subjunctive	present	*ka*(*m*)
Florina Arli	subjunctive	present	*ka*
Serbian Kalderaš	present-subjunctive	future	*ka*
Lovari	present-subjunctive	future	–
Rumungro	present-subjunctive	future	–
Roman	present-subjunctive	future	–
Sinti, Manuš	subjunctive	present-future	–
Finnish R	subjunctive	present-future	–
Russian R	present-subjunctive	present	*l-*
Latvian R	present-subjunctive	present-future	–
Welsh R	present-subjunctive	present-future	–

their function, often in connection with the introduction of an analytical future category. Table 4.3 shows the distribution in some dialects (see Chapter 13 for details).

Noteworthy is the geographical distribution of the developments: In the Balkans (Sepečides, Rumelian Romani, Kosovo Bugurdži, Florina Arli), the long forms are confined to the present indicative, and the future is expressed by a future particle (followed by the subjunctive). In central Europe (Lovari, Rumungro, and Roman), the short forms take over also a present indicative meaning, while the long form specialises for future. Serbian Kalderaš, a migrant dialect, shows contamination of the central European pattern with the Balkan pattern. The original state of affairs is preserved in the western, German–French and Scandinavian dialects. Elsewhere, combinations are found: an ongoing shift in the expression of the present indicative from long to short forms, combined with a loss of the future meaning of the long forms only through the introduction of an analytic future in Russian Romani.

4.3.3. Data collection procedures

The RMS database was designed initially on the basis of a core sample of recently published, exhaustive descriptions of Romani dialects. To these, data were added from a selection of texts, as well as additional more brief or older

descriptions. The variation depicted by this sample of sources covered the anticipated variation in the areas of morphological paradigms, while for syntax and syntactic functions typological references were consulted. Dialectological sources reached their limit at this point, for it proved extremely difficult to answer the syntactic-typological questions, including questions concerning categorisations such as the functional distribution of indefinites, based on published descriptive sources. A procedure now had to be designed that would allow the project to elicit structures in a way that would cater directly to the questions covered by the database. A questionnaire was then designed for this purpose. In its first edition from November 2001, the Romani Dialectological Questionnaire[5] contained a wordlist of 234 items, as an indication of lexical and phonological variation among the dialects; full conjugations of around 50 verbs, covering all inflection classes including likely borrowings; and some 750 short sentences, designed to cover systematically all relevant areas of morphosyntax.

The questionnaire was then used in elicitation sessions throughout Europe, by junior researchers and research assistants trained on the project, as well as by local student assistants of Romani background.[6] Using the respective majority language as the elicitation language (i.e. using translations of the English questionnaire), the fieldworkers asked native speakers to translate the words and sentences into their dialect of Romani. The five-hour sessions were recorded on tape, and the Romani translations transcribed. For the transcriptions, the teams used a table template in Excel prepared by the project. The table contains a column with tags, representing grammatical categories and functions. These can be fairly general, such as abbreviations for word classes ('PR' for Prepositions, 'DEM' for Demonstratives, 'NUM' for Numerals, 'IND' for Indefinites, and so on), or rather specific semantic-syntactic func-

Table 4.4. Example of elicited sentences with tags (Polish Romani)

111.	She came out of the house.	*joj vygeja khərestyr*	PR, CR-LOC-I-A
112.	I heard that other Roma live here as well.	*me šundžom kaj vavir roma bešte (a)daj*	CC , DEM, IND, UM, DEM-A, CC-E
113.	I saw a man walking down the street.	*me dikčom manušes syr našol pe ulica*	VI, ART, CC, CR-Temp, CR-LOC-V-C
114.	I have nothing left to give to you.	*man nan'i čhi so me tuke te dav*	IND, PP, CR-REC-G, CC-P
115.	There she is!	*joj doj sy*	PP, COP

Table 4.5. Example of modal constructions with 'want' (Lithuanian Romani)

I do not want to go to town.	*me na kamom te dž'ov do foro*	VI, NEG, CC, MOD, CR-LOC-I, CC-MO-W
Every evening he wanted to go somewhere.	*každy b'el'v'el'jov kamja varikajto te dž'al*	IND, CC, CR-Temp, CR-TEMP-SI-P, CC-MO-W
She didn't want anything to drink.	*joj na kamja niso te pijel*	IND, CC, CC-MO-W
She said to the witch: I want to be young again.	*joj pxend'a čovaxan'ake: me snova kamom te javov terny*	VI, CR, CR-REC-O, CC-MO-W

tions, for instance 'CR-LOC-I-A' for 'Case Representation: Incorporative Ablative' (as in *out of the house*), 'AC-TS-P' for 'Adverbial Clause: Time: Simultaneous: Punctual' (as in *just as . . .*), and so on. The tags closely mirror the divisions and categorisations within the database. Sorting the data columns according to the tags allows the user to access a sub-corpus of sentences that illustrate the particular word class or function in question.

Table 4.4 illustrates a typical data grid derived from the questionnaire, showing numbered sentences, the transliteration of the Romani translations provided by the speakers, and the adjoined list of tags (here: for the variety spoken by the Polska Roma of Pabianice near Łódż, Poland).

The sorting procedure is illustrated by a selection of sorted sentences in Table 4.5. Here, examples are given from a sub-sample created by searching the data grid for the Lithuanian Romani elicitation for the tag 'CC-MO-W' representing Complement Clauses: Modality: 'Want', or identical subject modal constructions with the matrix verb *want*. Quick retrieval of a sub-sample facilitates entry of data into the database, as well as just an informal impression of the structural features of any particular category of function.

The procedure allows easy retrieval also of comparative data, using the same sample sentences to evaluate structural differences among the dialects (Table 4.6).

4.4. Summary: Features and problems of the sample

Our sample of varieties of Romani is obviously not intended to be representative of the language diversity of the world. Nevertheless, its aim is to enable generalisations about variation, the relative frequency of phenomena, and preferences and dispreferences in structures. We pursue these aims by drawing

Table 4.6. Sample comparison of sentences in different dialects

Dialect	'I wanted to go home'	'She didn't want anything to drink'
Lithuanian R	*me kamjom te džov kxere*	*joj na kamja n'iso te p'ijel*
Klenovec Rumungro	*kamāhi te džan khēre*	*na kamlahi ňič te pijen*
Polish Romani	*me kamdžom te džał khəre*	*joj na kamełys čhi te pijeł*
Šumen Drindari	*me mangi tə žaa mange ando kher*	*oj na mangelas tə pel nikači*
Epiros R	*kamamas te džavas to kher*	*oj na kamelas čumune te pjelas*

on a database that captures the formal representation of (universal semantic) functions of language, as well as the representations of certain formal categories that are language specific (specific to Romani). While our sample cannot deliver any insights into universal linguistic diversity, we are able to draw on the unique advantages of a dialect sample: Assuming a common Early Romani ancestor whose structural features and time depth can be reconstructed, it allows us to investigate the degree of variation under controlled conditions ('controlled' in the sense of consistent and non-variable), and so the degree of variability resulting from change.

What makes our sample typological is its focus on the way in which universal functions of language happen to be represented formally in the varieties which we sample. Based on the kind of structural representations of functions and categories that are, potentially at least, universal, we divide the data into types, and attempt to find implicational relationships between the occurring types. What we are unable to do with this sample is offer an absolute statistical quantification of the data. This is not possible, partly because data are not always available on each and every structure from each of the sample varieties. Gaps – not even necessarily in the representation of entire categories, but just of individual values of those categories, in individual structures – may prevent complete comparability of the data. A further difficulty with quantification is connected to the specific character of the cross-dialectal sample: The cross-dialectal frequency of types is difficult to determine statistically, since the criteria for the independence of instances in a dialect sample are never entirely clear. Thus, separating the Erli variety of Sofia from the Yerli variety of Velingrad as two different 'dialects' and hence two independent instances within the sample would be based on just some isoglosses, disregarding others; the two varieties are spoken in separate locations, but within the same region, and

show a number of differences, but also a remarkable number of similarities. It is therefore never clear how dialects might be counted – whether a group of closely related dialects should constitute just one independent instance, or whether several speakers from the same community whose speech shows slight differences might be counted as different sources/dialects.

These problems are normally not encountered in a cross-linguistic survey where the data derive from published grammatical descriptions, each constituting a 'source' and therefore an 'instance' within the sample. They confront us however in the context of our elicitation work. On the whole, we take a pragmatic approach to this and the other sampling problems: Sources from different communities are treated as independent dialects, and sub-samples take into consideration those dialects for which relevant data is documented. It is for this reason that we avoid in most cases applying any quantitative methodology, and rely instead on the qualitative division into types, and the evaluation of relations among those types.

Chapter 5
Early Romani

In this chapter we provide a reconstructed outline of the forerunner of Romani dialects, which we will call *Early Romani*. Our purpose is to survey the structures that constituted the point of departure for the variation which we find today among the varieties of the language. Early Romani is not attested, and one must rely on dialect comparison as well as on comparisons with other Indo-Aryan languages in order to identify archaic structures among those that are still in existence, and postulate possible forerunner forms. The comparative method has been applied in such a way for Romani since the late-nineteenth century (cf. discussion in Matras 2002, Ch. 3), and has in recent years experienced a revival in the modern context of Romani dialectology and typology (cf. Matras, Bakker, and Kyuchukov 1997; Elšík and Matras 2000; Matras 2002). We refer to this work for the detailed arguments, and limit ourselves here to a presentation of what we consider to be reasonable outcomes of reconstruction scenarios.

We define *Early Romani* (ER) as the latest stage in the history of the language prior to the dispersion of Romani-speaking population groups throughout Europe and the consequent split into dialects and dialect families (see also Matras 2002). Despite the absence of any documentation or even attestation, we have both linguistic and extralinguistic clues concerning the time and location where ER was spoken. The earliest documentary evidence of an expansion of the Romani population into the northern Balkans, and beyond that, into central Europe, dates from the late-fourteenth century (cf. Fraser 1992 for a useful overview). A strong Greek element is shared by all dialects of Romani. It includes lexical vocabulary, grammatical vocabulary (function words, including numerals, indefinite pronouns, particles and adverbs of time and phase), derivational morphology, and patterns of syntactic typology that are most likely to have emerged in contact with Greek (such as the formation of relative clauses, a preposed definite article, verb-medial word order, and a split between factual and non-factual complementisers). The most striking evidence of Greek influence is, however, the incorporation of Greek nominal inflectional class markers (for subject-case nouns and adjectives) and of Greek verb inflection class markers (including tense-aspect marking, participial inflections, and marginally also subject concord inflections for the third-

person singular). All this points to the powerful influence of Greek, conditioned in all likelihood through a prolonged stay in Greek-speaking territory in Byzantium (see already Miklosich 1872–1880: III). This makes it convenient to define ER as the period during which Romani was in contact with Greek, prior to the decline of Byzantium and the beginning Romani migrations out of the Balkans and the subsequent split into isolated dialects.

As for the location of ER, one should bear in mind that Byzantium of the tenth-fourteenth century did not overlap with today's Greek-speaking area, but stretched as far as Anatolia and beyond; indeed Greek was spoken in Anatolia until the first decades of the twentieth century. This detail is sometimes overlooked in discussions of Romani history, but it could be crucial toward an understanding of the time line of Romani migrations westwards from India, and how this time line can be reconciled with linguistic evidence. Conventionally, the presence of Iranian and Armenian loans in Romani is taken as evidence of prolonged settlement periods in Iran and Armenia. For the latter, the present-day Armenian Republic north of the Ararat is often taken as the point of reference, and migration routes are pictured as having led north from Iran to Armenia, then south again along the Black Sea coast, to present-day Greece. But in tenth-century East Anatolia, Roma would have had speakers of Greek, Armenian, and Iranian languages (such as Kurdish, but possibly also Persian) as their neighbours. It may therefore be quite difficult to draw a clear-cut distinction between Romani in contact with western Asian languages, and Romani in contact with Greek, and there could well have been a period of overlap, and so possibly also a rapid, rather than gradual migration from India to Byzantium.

5.1. Lexicon

The borrowability of non-basic lexicon makes it impossible to estimate the actual number of pre-European lexical roots in ER. Adding together the pre-European vocabulary of present-day dialects, one might arrive at an estimate of possibly around 700 Indo-Aryan roots, with some 200 to 250 other pre-European roots, mainly of Iranian and Armenian origin. The fact that all dialects of Romani have Greek words for the numerals 7 to 9 (and often for 30 to 50 and above; see Chapter 11 for details) indicates that these were Greek borrowings in ER, and gives us a rough idea about the general extent of lexical borrowing from Greek in ER. It is likely that, for most domains except perhaps the most immediate, intimate lexicon (e.g. close kin, parts of the body,

very basic foods and animals, verbs of movement), there was free use of Greek words. The adoption of Greek inflection class morphology (see below) indicates that lexical words were employed with elements of their original Greek morphology. As contact with Greek was lost, through migration or changing linguistic orientation following transitions in power (from Greek to Turkish, in the Balkans), the dialects preserved some Greek lexicon, but began to replace much of it with borrowings from the new contact languages.

5.2. The sound system

ER consonant phonemes preserved the distinctive aspiration in voiceless stops, which was inherited from the northwest Indic ancestor language, and quite possibly also one or several (possibly allophonic) retroflex consonants.[1] It is likely that some of the Middle Indo-Aryan retroflex sounds had already shifted before the ER period, but at least a number of sounds, notably /ḍ/ as in /*ḍom/ > *řom*, *lom* 'Rom', /ṇḍ/ as in /*maṇḍa/ > *manřo*, *mando*, *manlo* 'bread' and /ṭṭ/ as in /*aṭṭa/ > *ařo* 'flour' appear to have survived well into the ER period. The variation in their continuation in present-day dialects might

Table 5.1. Early Romani consonant phonemes

	Bilabial	Labiodental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Glottal
Plosive: voiceless	p		t (t')				k		
Plosive: voiceless aspirated	ph		th				kh		
Plosive: voiced	b		d (d')		(ḍ)		g		
Nasal	m		n		(ṇ)		[ŋ]		
Affricate: voiceless			c	č					
Affricate: voiceless aspirated				čh					
Affricate: voiced			(dz)	dž					
Fricative: voiceless		f	s	š				x	h
Fricative: voiced		v	z	(ž)				(ř)	
Trill			r		(ṛ)				
Lateral			l		(ḷ)				
Approximant						j			

possibly suggest a set of retroflex variants, for which in turn the dialects each selected one non-retroflex counterpart, normally /ř/ [ʀ, r:], /r/, /n/, /d/ or /l/. In addition to consonants that were retained from the Indo-Aryan inventory, ER also had phonemes that had entered the language with Iranian and later with Greek borrowings: /v/, /f/, /z/, /c/ and possibly also /dz/ and /ž/. There appears to have been allophonic palatalisation of dental stops, and perhaps also of velars, before /i/, as word-specific palatalisation and affrication patterns are often found in the individual dialects (cf. *tikno* 'small' > *cikno*, *keti* 'how much' > *keci*, and more).

Other typical developments in the consonant inventory of the dialects following the ER period include palatalisation around jotated segments (e.g. /dj/ > /d' ď dź dž dz/), aspiration of /s/ > /h/ in grammatical inflections (in all likelihood inherited in some cases from MIA, then expanded), loss of final /s/, reductions of initial /a/, and prothesis of /j/, /v/ and sometimes other consonants, as well as prothesis of /a/.

The ER vowel system appears to have merged with that of (late) Greek, giving a five-vowel system /i e a o u/, with no opposition of length. Following the ER period, individual dialects have modified the vowel system. Later developments include the addition of the central vowels /ə/ and/or /y/, forward shift of stress (e.g. *čhavó* 'child' > *čhávo*), and the introduction of vowel length in some dialects (e.g. *čhāvo*).

5.3. Nominals

ER nouns had two genders, masculine and feminine, and two numbers, singular and plural. Modifiers generally agreed with their head nouns. Agreement was marked on adjectives, definite articles, demonstratives, and via a typical Indo-Aryan double-case phenomenon (see Plank 1995), also genitive attributes: *čhav-es-ker-o dad* 'the boy's father', *čhav-es-ker-i daj* 'the boy's mother'. Gender agreement was neutralised in the plural. Case marking appears to have been sensitive to animacy to some extent.

5.3.1. Case marking and declension classes

The case marking system preserved the late Indo-Aryan system of three layers (cf. Masica 1991). The most archaic layer, Layer I, is closest to the noun stem, distinguishes nominative and oblique cases, and is sensitive to declension

Table 5.2. Early Romani nominal declension classes

Class (abbreviation)	Example	NOM		OBL	
		SG	PL	SG	PL
OIKOCLITIC (PRE-EUROPEAN)					
C-masculines: NOM.PL *-a* (M0-a)	*kher* 'house'	–	*-a*	*-es-*	*-en-*
C-masculines: abstract (M0-A)	*čačipen* 'truth'	–	*-a*	*-as-*	*-en-*
C-masculines: NOM.PL – (M0−0)	*vast* 'hand'	–	–	*-es-*	*-en-*
o-masculines (Mo)	*šero* 'head'	*-o*	*-e*	*-es-*	*-en-*
i-masculines (Mi)	*pani* 'water'	*-i*	*-j-a*	*-j-es-*	*-j-en-*
C-feminines: unjotated (F0-U)	*džuv* 'louse'	–	*-a*	*-a-*	*-en-*
C-feminines: jotated (F0-J)	*suv* 'needle'	–	*-j-a*	*-j-a-*	*-j-en-*
i-feminines (Fi)	*piri* 'pot'	*-i*	*-j-a*	*-j-a-*	*-j-en-*
XENOCLITIC (EUROPEAN)					
o-masculines (*Mo)	*foros* 'town'	*-os*	*-i*	*-os-*	*-en-*
u-masculines (*Mu)	*papus* 'grandfather'	*-us*	*-i*	*-us-*	*-en-*
i-masculines (*Mi)	*sapunis* 'soap'	*-i*(*s*)	*-ja*	*-is-*	*-en-*
a-feminines (*Fa)	*cipa* 'skin'	*-a*	*-es?*	*-a-*	*-en-*

class (Table 5.2) (cf. Elšík 2000b for a reconstruction of ER declension classes). Among the inherited or *oikoclitic* declension classes (traditionally referred to in Romani linguistics as 'thematic') we find distinct classes for the masculine and the feminine, further differentiated into several vocalic and consonantal (sub)classes and, within the feminine, also into jotated and unjotated (sub)classes; there was also a distinct declension (sub)class for abstract nominalisations. The inflections tend to derive from OIA/MIA nominal derivational markers in the nominative, and from remnants of the OIA/MIA genitive case in the oblique.

In addition to the inherited declension classes, ER develops several classes for borrowed nouns, or *xenoclitic* classes (traditionally referred to in Romani linguistics as 'athematic'). The formation of these declension classes is based on the incorporation of Greek nominative inflections, whereby the neuter gender is integrated into the masculine classes, resulting in variation in the presence of the (originally masculine) inflection *-s* in the relevant classes. In the oblique, the vowel of the oikoclitic singular masculine marker has been assimilated to the vowel of the Greek-derived nominative markers (e.g. **-es-* > *-os-* in the xenoclitic *o*-masculines). The plural oblique of xenoclitic classes was

not distinct from that of oikoclitic classes in ER. This Greek-based class differentiation became a stable component of the ER declension system. Occasional loans from languages other than Greek were assigned one of the Greek-derived inflections. As the Romani-speaking population groups broke away from Greek-speaking territory, the Greek-derived system continued to serve as a grid for the incorporation of more recent loans from the new contact languages. The system eventually underwent considerable levelling in the individual dialects, but continues in present-day dialects to mark out European loanwords.

The nominative case was that of the subject as well as of the inanimate direct object. The oblique case had several uses, which included the marking of the animate direct object, and quite likely also the marking of the possessor, the external possessor (as in 'I am cold' or 'I have a headache'; see Koptjevskaja-Tamm 2000, König and Haspelmath 1997), and the recipient of the verb 'to give'. Apart from these functions, the oblique served as the basis for Layer II case markers. These derive from a set of enclitic markers which in ER (and probably earlier) are fused with the noun, forming an agglutinating set which shows consonant alternation determined by the final segment of the preceding preceding Layer I suffix (see Table 5.3).

Table 5.3. Early Romani Layer II case markers (Sg/Pl)

Dative	Locative	Ablative	Sociative	Genitive
-ke/-ge	*-te/-de*	*-tar/-dar*	*-sa*	*-ker-/-ger-*

Layer II suffixes mark a Dative, Locative, Ablative, Sociative (Instrumental-Comitative), and a Genitive which agrees with its head through Suffixaufnahme. The Locative in *-te/-de* appears to derive from the older dative, at the expense of an old locative in **-me*, which may have existed earlier (as in other NIA languages, including Domari), while the new dative in *-ke/-ge* is essentially a benefactive.

5.3.2. Adjectival modifiers

There are several distinct patterns of modifier inflection in ER (Table 5.4). The distinction between oikoclitic and xenoclitic forms was found in (descrip-

Table 5.4. Early Romani adjectival inflection

Class	SG				PL	
	NOM.M	NOM.F	OBL.M	OBL.F	NOM	OBL
Oikoclitic	*-o*	*-i*	*-e*	*-a*	*-e*	
Xenoclitic	*-o*		*-on-e*	*-on-a*	*-a*	*-on-e*
Demonstrative	*-va*	*-ja*	*-l-e*	*-l-a*	*-la* (*-na*)	*-l-e*
Article	*o*(*v*)	*e*	*ol-e*	*ol-a*	*ol*	*ol-e*

tive) adjectives as well. Oikoclitic forms were typical of inherited descriptive adjectives, but also of participles, of the interrogative indefinite *sav-o* 'which', of the deictic determiner *asav-o* 'such', and of the universal determiner *savoř-o* 'all, entire, the whole'. Cardinal numerals and the quantifier *but* 'much, many' lacked overt nominative inflections, but took oblique suffixes. Xenoclitic adjectival inflections derive in part from Greek (in the nominative), or show the intrusive suffix *-on-* plus oikoclitic inflections (in the oblique). They are assigned to Greek-derived adjectives (and, as with nouns, they later accompany adjectival loans from subsequent contact languages in the individual dialects), as well as to the ordinal numerals (which are formed through addition of Greek-derived *-to*) and the Slavic-derived quantifier *vsako* 'every'. Marginally, there were indeclinable adjectives (*šukar* 'beautiful, nice'), quantity interrogatives (*keti* 'how much'), or quantity deictics (*ati* 'as much as').

Different patterns of agreement characterised two other modifer classes: demonstratives and definite articles. The nominative singular demonstrative inflections M *-va* and F *-ja* may have derived from an assimilation of earlier, regular adjectival inflections with a postposed deictic **-a* (cf. Matras 2002: 107–108), which together replaced the original consonantal deictic root. In the oblique singular, and in the plural forms, the original deictic root *-l-* is preserved. In the nominative plural, the root is followed by the very same postposed deictic *-a*, while the oblique forms show inflections that match the oikoclitic inflection pattern of adjectives. The definite article is related to the demonstrative, and we find similar forms in the more conservative oblique inflections. In the nominative, the forms were shortened, giving a new pattern of inflection. The simplification of the nominative singular forms in particular may have been influenced by the corresponding Greek form (*o*, *i*).

5.3.3. Demonstratives and related forms

Here we consider the formation of demonstrative roots, and related expressions. The original deictic root *-l-* tends to be preserved in the oblique forms and in the nominative plural. The pronominal demonstrative took nominal inflections in the oblique, while in the nominative we find the adjectival-demonstrative inflections discussed above (see Table 5.4). The outstanding feature of ER demonstratives is their renewed composition: new consonantal roots are preposed to the older demonstratives. These consonantal roots derive from the deictic expressions of location, *adaj* 'here', *odoj* 'there', *akaj* 'precisely here', *okoj* 'precisely there'. The roots in *-d-* thus indicate general deictic reference, while those in *-k-* indicate specificity. The vocalic patterns – *-a-* versus *-o-* – indicate distance or source of reference (situation vs. discourse). The renewed paradigm constitutes a four-term system. This same system is continued in the more conservative dialects (in the Balkans, in some of the Italian dialects, and in Welsh Romani). In other dialects, some of the forms are simplified to *kava*, *dava* etc., while in others various forms of 'strengthening' occur (e.g. **aka-adava* > *kadava*).

There are two competing forms of the third-person pronoun, both deriving ultimately from demonstratives. The ancient demonstrative stem in *-l-* is continued, with normal nominal inflection (of the oikoclitic vowel classes), in a reduced pronoun, which presumably was reserved for highly continuous referents: *lo*, *li*, *le*. Later on, many of the dialects lost this form completely, or restricted its distribution to non-verbal predications. Alongside this older set,

Table 5.5. Early Romani deictic and anaphoric expressions

	Nominative			Oblique		
	SG.M	SG.F	PL	SG.M	SG.F	PL
Demonstratives:						
proximate plain	*adava*	*adaja*	*adala*	*adales*	*adala*	*adalen*
proximate specific	*akava*	*akaja*	*akala*	*akales*	*akala*	*akalen*
remote plain	*odova*	*odoja*	*odola*	*odoles*	*odola*	*odolen*
remote specific	*okova*	*okoja*	*okola*	*okoles*	*okola*	*okolen*
Third-person pronoun	*ov* (*av*)	*oj* (*aj*)	*ol,on* (**al*)	(*o*)*les*	(*o*)*la*	(*o*)*len*
	lo	*li*	*le*			
Definite article:	**ov* > *o*	**oj* > *e*	*ol*	(*o*)*le*	(*o*)*la*	(*o*)*le*

ER had a renewed set of emphatic subject pronouns: *ov*/*av*, *oj*/*aj*, *ol*/*al*. These derived from the demonstrative in its intermediate stage – following the loss, in the nominative singular, of the ancient stem in *-l-*, but prior to the 'strengthening' which characterises the emergence of a new four-term system in ER – in a syncopated form, showing loss of the final vowel. It is from the remote forms of this paradigm – *ov*, *oj*, *ol* – that the definite article emerged, copying the pattern of a preposed definite article in the contact language, Greek.

5.3.4. Personal pronouns

Third-person pronouns were dealt with above, and we limit ourselves in this section to deictic personal pronouns, and reflexives. ER inherited an archaic system of first and second-person pronouns (see Table 5.6). They show numerous irregularities, such as the marking of the plural by the suffix *-m-* in 1PL *a-m-* and 2PL *tu-m-*. The 1SG pronoun stands out in having distinct suffixes in the nominative and oblique, respectively, while in the other pronouns the nominative and oblique are identical. The direct object (accusative) case is the oblique with no additional inflections, save with the 2SG pronoun, where there is a distinct accusative marker *-t*.

Whereas third-person pronouns formed their possessive (genitive) forms in the same fashion as nouns, i.e. with the genitive marker suffixed to the oblique stem (e.g. 3SG.M oblique *les-*, genitive *les-ker-o* 'his'), first- and second-person pronouns showed several forms. The plurals contained the possessive marker *-ar-*, which was suffixed directly to the base. The 2SG had the possessive suffix *-ir-*, while in the 1SG the suffix appears to have been *-inř-* (cf. Elšík 2000a). The dialects continue the plural forms, but show various developments in the singular possessive pronouns (see also Chapter 7).

The Early Romani reflexive *pes-* was used in coreference with third-person antecedents of both numbers. It is not clear from the current cross-dialectal

Table 5.6. Early Romani first and second-person pronouns

	1SG	2SG	1PL	2PL
Nominative	*m-e*	*t-u*	*am-en*	*tum-en*
Oblique stem	*m-an-*	*t-u-*	*am-en-*	*tum-en-*
Possessive stem	*m-inř-*	*t-ir-*	*am-ar-*	*tum-ar-*

distribution whether the Early Romani possessive stem was the regular **pes-ker-* or the irregular **p-inř-* (see Chapter 7 for details).

5.3.5. Interrogatives

Early Romani interrogatives are shown in Table 5.7. Cause and goal were probably not distinguished. It is not clear whether there was a distinct size interrogative (possibly **ki-bor*; see also Chapter 20).

Table 5.7. Early Romani interrogatives

Value	Early Romani	English
Determiner	*savo*	'which, what sort of'
Person	*kon*	'who'
Thing	*so*	'what'
Place: stative/directive	*kaj*	'where, whither'
Place: separative	*katar*	'whence'
Time	*kana*	'when'
Manner	*sar*	'how'
Cause /goal	*soske*	'why'
Quantity	*keti*	'how many/much'

5.3.6. Indefinites

Three indefiniteness series may be reconstructed for Early Romani (see also Chapter 19): a specific-to-negative *kaj*-series (covering a wide range of indefiniteness meanings, probably from specific via irrealis and negative polarity to negative proper), a free-choice *moni*-series, and a universal series. Reconstructable forms of the *kaj-* and *moni*-series are given in Table 5.8 (see Elšík 2000c).

Table 5.8. Early Romani indefinites

Value	Specific-to-negative	Free-choice
Determiner	*kaj* (*daj*)	*moni*
Person	*kon-jekh*, *kaj-jekh* (*daj-jekh*)	*ko*(*n*)*-moni*
Thing	*či*, *kaj-či*, *kaj-ni-či*	*či-moni*
Place	*kaj-ni*	*kaj-moni*

The *kaj*-indefinites consist of the following components: the inherited determiner *kaj*, rarely *daj*; the thing indefinite *či*, possibly a loan from Iranian; interrogative bases (*kon* 'who' and *kaj* 'where'); the numeral *jekh* 'one'; and the assumed focus particle **ni* (cf. the attested *vi*, *li* 'also'). The *moni*-indefinites, apart from the determiner, are derived from interrogatives or specific-to-negative indefinites. The origin of the free-choice suffix *-moni* is obscure. It may be a result of internal grammaticalisation of the assumed focus particle **moni* 'only', possibly consisting of *mo*(*no*) 'only' of Greek origin and the focus particle **-ni*. Indigenous universal quantifiers include the particle *sa* 'everything, all, always' and the adjectival *savořo* 'all'.

5.4. Verbs

5.4.1. Valency and loan verb integration

ER inherited a set of valency-changing morphemes, reminiscent of MIA verb morphology. Two valency-increasing markers, *-av-* and *-ar-*, were old, and their productivity appears to have been restricted. Nonetheless, they continued to play a role in the formation of causatives and in the formation of transitive verbs from non-verbs (adjectives, participles, and to some extent also nouns). Alongside these two markers there was a younger valency-increasing morpheme, *-ker-*, a grammaticalisation of the verb *ker-* 'to do'. This marker played a slightly more active role in word-formation, especially in deriving transitive verbs from nouns. It appears that *-ker-* also combined with the two older morphemes as a means of strengthening their valency structure (*-av-ker-*, *-ar-ker-*). Transitive word-formation also relied on another verb, *d-* 'to give', which however seems to have been confined to lexical derivations, and did not have a grammaticalised valency-increasing function. The common valency-decreasing morpheme was *-jov-*, derived originally from the verb *ov-* 'to become'. It was used productively with transitive participles to derive middles or 'mediopassives' (*kerdo* 'done', *kerdjov-* 'to be done'), and with adjectives and some nouns to derive inchoatives. Another marker, *-áv-*, emerged in similar function, and appears to have enjoyed even greater productivity that *-jov-* in ER, although variation seems quite likely, and dialects differ in their retention and distribution of the two. This later marker derives from the grammaticalised verb *av-* 'to come', which appears to have been used at least as a variant in similar functions to *ov-* 'to become'. Though internal grammaticalisation is a possibility, contact influence from Kurdish *hatin* 'to come', which

is also used as an intransitive auxiliary, may have played a role in the late Proto-Romani period, immediately preceding ER.

Valency markers appeared to have played a role in the integration patterns of loan verbs. In ER, there appears to have been a free licence to integrate Greek lexicon, including verbs, with the exception perhaps of a limited range of semantic domains. Greek verbs will have been adopted with their inflection class markers, such as *-ín-*, *-íz-*, etc., which in turn also marked tense (present vs. aorist). Thus, Greek-derived verbs continued to carry Greek-derived tense inflection when used in Romani (*jiríz-* 'to go', *jirís-* 'to have gone'; *graf-* 'to write', *graps-* 'to have written', etc.). The Greek stem with inflection class marker was then followed by an internal, Romani 'verbaliser', marking out valency and so assigning the borrowed stem its status as a verb within the recipient language. This system could have been inherited from the older Indo-Aryan ancestor language, or it could simply have emerged in congruence with a strategy for adapting loan verbs that was, and still is, common in the western Asian area (cf. Turkic, Iranian, and Indo-Aryan languages) (see discussion in Matras 2002: 128–135). The markers involved were the transitive markers *-av-*, *-ar-*, and *-ker-*, with transitive verbs, and *-áv-* with intransitive verbs. These were then followed by the regular Romani tense–aspect–mood and concord markers, so that the entire, complex strategy of loan verb inflection constituted in effect a derivational strategy, even though it was sensitive to Greek inflection class membership. As with nouns, this system of integrating loan verbs by respecting and marking out their Greek-based inflection class affiliation became a stable component of the language. As contacts extended beyond Greek, loans verbs from the new contact languages were initially incorporated into the same, Greek-oriented system. With the declining number of Greek verbal roots, and the increasing number of loan verbs from other contact languages following the dispersion of the dialects, the assignment of Greek-based inflection class became, of course, dysfunctional. The outcome was a series of dialect-specific levelling processes, applied to the full paradigm of ER loan verb adaptation marking (see Chapter 15 for details). In effect, one or several loan verb markers were generalised. These may contain, in individual dialects, just a Greek-derived morpheme (e.g. *-in-*), or just a valency morpheme (e.g. *-ar-*), or traces of both (e.g. *-is-ar-*, *-is-ker-*). Valency distinctions are consequently no longer strictly observed in the dialects, either, though the original ER past-tense intransitive marker *-is-ájl-* from **-is-áv-il-* continues to function, in many dialects, as a specific past tense formation for intransitive loans. In some dialects, this adaptation system is also applied to new formations within the pre-European component, such as iterative *-in-ker-* or inchoative *-is-aj-*.

5.4.2. Inflection classes

ER had distinct inflection classes for non-perfective and perfective (sometimes called past, aorist, or preterite) stems. In the non-perfective stems, there were two classes, a consonantal class (*ker-* 'to do'), which included the majority of verbs, and a rather marginal vocalic class (*xa-* 'to eat'). The difference between the two non-perfective classes is merely in the shape of the vowel that connects the stem to the subject concord marker (see Section 5.4.3).

The class division among perfective stems was somewhat more complex. Here, the differences were in the form of the perfective marker which attached to the root to form the perfective stem. These perfective markers all derive from an inventory of OIA/MIA participial and adjectival inflections, a development which goes back to the loss of the old past-tense formation and the generalisation in early NIA of participles instead. The most common marker in Romani is the historical **-t-*, which took on various forms in different phonological environments. It appears that the original *-t-* was retained in ER with roots in voiceless consonants, and to some extent with roots in */m/*. The latter had already begun to move into a different class, namely that in *-l-*, and this process continued in most dialects with the roots in voiceless consonants, to different extents. The same historical participle marker was voiced to *-d-* following roots in voiced consonants, and shifted to a dental lateral *-l-* following vowels.

A different participial marker, *-in-*, was employed for the perfective stem of the monoconsonantal roots *d-* 'to give', *l-* 'to take', and *s-*/*h-* 'to be'. Yet another inflection, the originally adjectival *-il-*, specialised in the semantically demarcated class of middle verbs, attaching normally to transitive participles (*kerd-il-o* 'was done'), and motion verbs (*av-* 'to come', *ačh-* 'to stay'). Final-

Table 5.9. Early Romani perfective inflection classes

Class	Perfective marker
Roots in (*m*) *k kh t č š s*	*-t-*
Roots in *n r l v*	*-d-*
Roots in vowels, (*m*)	*-l-*
Monoconsonantal roots *d-*, *l-*, *s-*/*h-*	*-in-*
Middle and motion verbs	*-il-*
Psych verbs in *-a*	*-n-*(*d*)-(*il*)-
Irregulars: *mer-* 'die', *dža-* 'go' etc.	*mu-l-*, *ge-l-*

ly, a small class comprising psych verbs in *-a* (*dara-* 'to fear') appears to have had its own, complex marking, evidently a combination of various perfective markers and adjectival extensions, quite possibly in variation. ER also inherited some irregular formations from OIA/MIA, including suppletion in *dža-* ~ *ge-l-* 'to go' and *ov-* ~ *u-l-* 'to become'.

5.4.3. Concord markers

The original OIA set of subject concord markers is nicely preserved in the ER non-perfective conjugation (see Table 5.10), probably more so than in most known NIA languages with the exception of Domari. Full concord markers appear in the class of consonantal stems (e.g. *ker-el-a* 's/he does'), while the vowel component disappears following vocalic stems (e.g. *xa-l-a* 's/he eats'). With Greek loan verbs, it seems that ER used, possibly alongside the corresponding indigenous marker, also the Greek-derived third-person singular marker *-i*.

The original OIA past conjugation having been lost, a new set of perfective (primarily past-tense) concord markers emerge, drawing originally on enclitic oblique pronouns, which are attached to the past participle by means of a jotated particle. In ER, the system is already well-formed and characterised by the jotation of the concord markers, which is later to give rise to various process of palatalisation, affrication, umlaut or de-jotation in the dialects. Exceptions to the jotation are the two cases where the participle is followed not by a person marker deriving from an original enclitic pronoun, but by an adjectival

Table 5.10. Early Romani subject concord markers

	Non-perfective		Perfective	
	Consonantal	Vocalic	Transitive	Intransitive
1SG	*-av*	*-v*	*-jom*	*-jom*
2SG	*-es*	*-s*	*-jal*, *-jan*	*-jal*, *-jan*
3SG	*-el*	*-l*	*-jas*	*-o* ~ *-i*
1PL	*-as*	*-s*	*-jam*	*-jam*
2PL	*-en*	*-n*	*-jan*	*-jan*
3PL	*-en*	*-n*	*-(in)e*	*-(in)e*

participle marker: These are the third-person plural and singular of intransitive verbs. The latter results in effect in a split between two series of perfective concord markers.

Recently documented formations from the very conservative Romani variety of Epiros in Greece (Matras 2004) suggests that ER may have also had object concord markers, which show agreement for gender and number, of the type: **dikht-jas-os* 's/he saw him', **dikht-jas-i* 's/he saw her', **dikht-jas-e* 's/he saw them'. Although late cliticisation of third-person object pronouns in Epiros Romani cannot be ruled out entirely, the pattern of object concord is strongly reminiscent both of that found in Domari (Matras 1999b), and of the formations found in the extreme northwestern or Dardic languages of India. Both these and Domari also show the same pattern of historical emergence of past-tense subject concord markers from enclitic pronouns, and it seems reasonable to assume that ER had inherited the complete package. This assumption is strengthened by the position of the object pronoun (*o*)*les* in most conservative present-day dialects after the verb (e.g. *dikhav les* 'I see him'). If ER had had a free-standing object pronoun, it seems likely that its position would have followed the Greek model, and it would have been placed before the verb (cf. Greek *ton vlépo*). This, since ER generally follows Greek word-order patterns, with the only exception of the positioning of the genitive attribute, which in Romani agrees with its head (see Section 5.6). We might therefore conclude that the present-day object pronoun derives from an ER demonstrative object; this is easily reconcilable with the form *oles*, which is that of an old demonstrative, and the postpositioning of a demonstrative object will have been fully compatible with Greek word-order patterns (cf. *vlépo aftón*). From this, it is possible to derive the conclusion that ER anaphoric (non-focused) object pronouns were enclitic, and were expressed as a second set of concord markers on the verb, as we still find today in Epiros Romani.

5.4.4. Tense, aspect and modality

The principal opposition within the ER system of TAM is that between the non-perfective event, characterised through the absence of explicit completion, and the perfective event, which is marked for completion (see Table 5.11). This opposition may be regarded as an aspectual one, though traditionally Romani grammars refer to the non-perfective Present/Future, and to the perfective Past (Preterite, or Aorist), as 'tenses'. The Present/Future with its forms in *-a* (e.g. *ker-av-a* 'I do/shall do') constituted the 'default' non-perfec-

Table 5.11. TAM categories in Early Romani

	Non-perfective root + NON-PFV concord	Perfective root + PFV marker + PFV concord
Non-remote	Present/Future	Past (Preterite, Aorist)
Remote	Imperfect	Pluperfect/Counterfactual
Intentional	Subjunctive	

tive category, unmarked for (deictic) tense, while the counterpart default perfective category (e.g. *ker-d-jom* 'I did') might be referred to as Past, Preterite or Aorist, but is similarly unmarked for actual tense, and merely marked for completion, though completion tends to overlap with past events.

Tense in the actual, deictic sense was expressed by the addition of a remoteness affix **-asi* to the Present/Future, to form the non-perfective remote, or Imperfect, or to the Past to form the perfective remote, or Pluperfect/Counterfactual. A syncopated form of the Present/Future, lacking the inflection *-a* (e.g. *ker-av*), was employed as a subjunctive, the only overtly marked modality category.

5.5. Other categories

5.5.1. Local adverbs

ER local adverbs encoded two orientation values: the stative-directive and the separative-perlative. Separative-perlative adverbs were derived from stative-directive adverbs by the old ablative suffix *-al* (e.g. *andr-al* 'from inside, through inside' < *andr-e* 'inside, inward', *dur-al* 'from far' < *dur* 'far', *kher-al* 'from home' < *kher-e* 'at home, home').

5.5.2. Prepositions

ER local prepositions are reconstructed in Fig 17.2. in Chapter 17. There were no distinct temporal prepositions. Non-local prepositions are: the privative *bi* 'without', the benefactive *vaš* 'for', and the causal *astjal* 'for'. The expressions *dži* 'up to, until' and *sar* 'as, than' (< 'how') were particles rather than prepositions in ER.

5.6. Syntax

ER was unique among the NIA languages in shifting to a verb-medial typology. Constituent order in the verb phrase will have been flexible, but with the dominance of SVO and VSO patterns which we still find today in most Romani dialects of the Balkans and central Europe. This shift in word order will have involved the prepositioning of local relation expressions and the emergence of prepositions in the language. The order of elements in the noun phrase will have remained more or less intact, as it did not conflict with the patterns found in the contact language, Greek: adjectival modifiers (adjectives, numerals, demonstratives) were preposed. ER maintained however the prepositioning of the genitive attribute, an Indic legacy (compared with the postposed genitive in Greek). As in the other NIA languages, the genitive attribute continued to agree with its head. The most striking development in the noun phrase, once again making ER unique among the NIA languages, was the emergence of the preposed definite article, copying the Greek pattern, but drawing on remote (anaphoric) pronouns.

In syntax, the most notable development was the retreat of most non-finite verb forms and the reliance on finite forms and conjunctions for clause combining. It is impossible to tell what exactly the Proto-Romani legacy was, and whether converbs of the type still found today in NIA had developed in the ancestor language, or were lost before the ER period. ER certainly had at least two gerundial constructions, one in *-indo*, possibly reinforced by the Greek model, and one in *-i*. But most adverbial subordinations and other strategies of clause combining seemed to rely on conjunctions, taken most frequently from the inventory of interrogatives. This could have been an earlier development, triggered through contact with Iranian, much like the reduction of the modal infinitive and its replacement by subject agreement in a finite, subjunctive complement clause (lit. 'I want that I go'). Perhaps the clearest piece of evidence in favour of syntactic convergence between ER and Greek in this domain is the emergence in ER of *kaj*, originally 'where', as a general subordinator and relativiser, and the emergence of a split between factual complementisers, for which *kaj* was used, and non-factual or subjunctive complementisers, for which the conditional particle *te* (originally probably a correlative particle) was employed (cf. Matras 2002: 179–185).

Chapter 6
Number

The category of number is, in Romani, coded in nominals (nouns, some pronouns, and adjectivals) and in verbs. It has two values: the singular and the plural. The search for number asymmetries is complicated by the fact that number is rarely expressed separately. Mostly, it cumulates with other inflectional categories: with person and aspect in verbs; with case and partly gender in nouns; and with case and gender in adjectivals. Thus many of the emerging complex hierarchies are ambiguous or difficult to evaluate, and many of them will be discussed also in sections on other categories.

The singular is the value that is more differentiated and exposed, tends to be extended in analogical change, and shows extended distribution. The plural, on the other hand, is the value that is more structurally complex, more diverse, and more likely to be borrowed. Although there are a few exceptions to these generalisations, the overall asymmetry with regard to our criteria is pronounced and consistent. There appears to be no salient erosion asymmetry.

6.1. Complexity

The singular tends to be less complex structurally than the plural. In this section, we describe instances of zero marking of the singular in the inflection of all three major morphological classes (verbs, substantivals, and adjectivals), and other instances of a lesser complexity of the singular in pronominal morphology and verb inflection. There are no instances of a lesser complexity of the plural. Nevertheless, the complexity asymmetry between the number values is never unconditional. In other words, the singular can be shown to be less complex than the plural only in specific paradigmatic contexts defined by cross-cutting categories: TAM and/or person with verbs, and case with nominals.

In **verbs** the form without an overt inflectional marker is, in Early Romani and in all dialects, the (second person) singular imperative of most verb classes, while the corresponding plural form contains an overt person–number marker (e.g. *ker* 'make!' vs. *ker-en* 'make.PL!').

Zero marking in **substantivals** is not as lexically general as with verbs: only nouns of the unproductive consonantal classes and some pronouns

(e.g. the interrogative *so* 'what', and in some dialects also *kon* 'who') ever occur without an overt inflectional marker. In most dialects, the only inflectional form that can be markerless is the nominative singular,[1] while the corresponding plural form contains an overt case–number marker (e.g. *vast* 'hand' vs. *vast-a* 'hands', or *daj* 'mother' vs. *daj-a* 'mothers'). In Early Romani and some dialects, the markerless form occurs in the nominative of both numbers with some consonantal nouns (e.g. *vast* 'hand, hands'), but only in the nominative singular with other nouns (e.g. *daj* 'mother' vs. *daj-a* 'mothers'). In any case, zero marking of the nominative plural implies zero marking of the nominative singular.

Though there are no markerless forms in **personal pronouns**, there is a clear asymmetry in structural complexity of the singular and the plural pronouns (Table 6.1). In the second person, the plural pronoun contains a more complex variant of the second person root: SG *t-* > PL *t*(-)*u-* (see analysis in Elšík 2000a). In the first person, the pronouns are strongly suppletive (cf. first-person singular *m-* vs first-person plural *a-*). In both persons, the plural is moreover marked by a separatist number suffix *-m-*; there is no such overt marking in the singular pronouns.

In Early Romani and in most dialects, there is no difference in complexity between singular and plural third-person pronouns. The nominative forms consist of a uniform root and an irregular case–number inflection that also cumulates gender in the singular (cf. Early Romani **o-v* 'he', **o-j* 'she', and **o-l* ~ **o-n* 'they'). Likewise, the oblique stems consist of a uniform root and regular substantival inflections of equal complexity (cf. Early Romani **l-es-* 'him', **l-a-* 'her', **l-en-* 'them'). Nevertheless, a few dialects have increased the complexity of the nominative plural forms through suffixing additional plural markers: either borrowed ones (e.g. Nógrád Rumungro **ón* > *ón-k* 'they'), or internal extensions (e.g. Northeastern **jon* > *jon-e* 'they'). Thus dialect specific developments in third-person pronouns tend to a greater complexity of the plural.

The criterion of zero marking with **adjectivals** is only relevant in some dialects. The majority of adjectives, those of the oikoclitic and xenoclitic vocalic

Table 6.1. Roots of personal pronouns

	SG	PL
First person	*m-*	*a-m-*
Second person	*t-*	*t*(-)*u-m-*

Table 6.2. Inflection of consonantal adjectives

	NOM.SG	NOM.PL	OBL
Type A	*-0*	*-0*	*-e* (~ *-a*)
Type B	*-0*	*-a*	*-0*
Type C	*-0*	*-a*	*-e* (~ *-a*)

classes, contain an overt inflectional marker in all of their forms, and so they are irrelevant here. The class of consonantal adjectives, which comprised a handful of underived adjectives (e.g. *aver* 'other', *xor* 'deep', *kuč* 'expensive', *midžak* 'evil', *šukar* 'beautiful', *tang* 'narrow') as well as the synthetic comparatives in *-eder*, was indeclinable in Early Romani, and continues to be so in a number of dialects (e.g. in Welsh Romani, Sinti, most Northeastern dialects, some Central, many Balkan, and some Vlax dialects). However, in other dialects some or all members of the class became declinable, due to extension of inflections from other nominals (Table 6.2).

In Bohemian and East Slovak Romani, Austrian Lovari, Kalderaš, Xoraxane, Kalburdžu, and variantly in Erli and Sepečides, inflections of the oikoclitic vocalic class of adjectives have extended into the oblique forms of the consonantal adjectives (Type A). In Šóka Rumungro, the consonantal adjectives took over the nominative plural suffix *-a* of consonantal nouns (Type B). Finally, both extensions have taken place in eastern Rumungro, Drindari, and variantly in Prilep (Type C). Now, only the nominative singular remained markerless in all dialects that have been affected by any of the two sorts of extensions. Through the presence of overt inflections in other categories the absence of any inflection became a relevant factor of structural complexity. Type B and C paradigms show that, in the nominative, the singular became zero-marked in comparison with the plural (the Type A paradigm is not relevant for number asymmetry).

Lesser structural complexity of the singular is also shown in the formation of third-person pluperfect forms in a few dialects. The pluperfect is generally formed by suffixation of the remoteness marker to corresponding preterite forms: e.g. Early Romani **kerdjom* 'I did' > **kerdjom-asi* 'I had/would have done'. However, the pluperfect forms derived from the third-person active participle preterite forms contained, on a synchronic analysis, an intrusive morpheme between the preterite form and the remoteness marker: e.g. Early Romani **kerde* 'they did' > **kerde-s-asi* 'they had/would have done'. Thus in Early Romani, the asymmetry held between the more complex participle-

Table 6.3. Intrusion in third-person pluperfect inflections

Dialect	3SG (FINITE)		3SG (PARTICIPLE)		3PL (PARTICIPLE)	
	PRET	PLPF	PRET	PLPF	PRET	PLPF
Early Romani	**-as*	**-as-asi*	**-o*	**-o-**s**-asi*	**-e*	**-e-**s**-asi*
Ajia Varvara	*-as*	*-as-as*	*-o*	*-o-**s**-as*	*-e*	*-e-**s**-as*
West Slovak	*-as*	*-ah-as*	(lost)		*-e*	*-e-**n**-as*
Latvian	*-a*	*-a-**s**-is*	(lost)		*-e*	*-e-**s**-is*
Šóka Rumungro	*-a*	*-ā-hi*	(lost)		*-e*	*-ē-hi*

based forms and the less complex finite forms, and was not indicative of a number hierarchy (Table 6.3).

The Early Romani asymmetry is continued in a number of dialects (represented by Ajia Varvara in Table 6.3). However, the singular participle form has been lost in many dialects, and so in those that retained the Early Romani formation of the pluperfect, the asymmetry became one of number: the third-person plural form contains an intrusion, while the only remaining third-person singular form does not. This is the case in West Slovak (shown in Table 6.3) and Welsh Romani. Nevertheless, the loss of the singular participle form does *not* necessarily lead to asymmetrical complexity in number. For example, in the Northeastern dialects (represented by Latvian Romani in Table 6.3) the intrusion now occurs in forms of both numbers due to a morphological reanalysis.[2] In most Rumungro dialects (represented by Šóka), on the other hand, there is no intrusion in either number, also due to a morphological reanalysis.[3] It may be concluded that reanalysis in the third-person pluperfect forms has favoured parallelism rather than asymmetry between the encoding of the two numbers, and that the number asymmetry in West Slovak and Welsh Romani appears to be a mere by-product of a different development, viz. the loss of the active participle.

6.2. Erosion

There are conflicting erosion asymmetries in the category of number. For example, the erosion hierarchy in middle verbs (which will be discussed in Chapter 7) is 3SG > 3PL > 2PL > 2SG > 1SG > 1PL: the singular is more likely to erode in the 3rd and first persons, but the plural is more likely to erode in the second person.

Table 6.4. Differentiation asymmetries in the category of number

	Person	TAM	Case	Gender	Class
Verbs	SG > PL	SG > PL	–	SG > PL	SG > PL
Nouns	–	–	PL > SG	(class)	none
Personal pronouns	SG > PL	–	SG > PL	SG > PL	–
Adjectivals	–	–	SG > PL	SG > PL	none

6.3. Differentiation

The singular shows, on the whole, a greater differentiation in cross-cutting categories than the plural (Table 6.4). In verbs there are more person, TAM, and gender distinctions, and in personal pronouns and adjectivals there are more case and gender distinctions. Verbs also possess more class distinctions in the singular, while with nouns and adjectivals, the asymmetries in class differentiation are inconclusive for the category of number. Case differentiation in nouns is exceptional in being less in the singular than in the plural. In this section, we first review instances of asymmetrical differentiation of number by cross-cutting categories, then go on to class differentiation, and at the end we discuss the criterion of exposition.

Early Romani non-perfective verb forms did not make a distinction between the second and the third **person** in the plural, while all persons were distinct in the singular. Thus, for each non-perfective set, there were three forms in the singular, and only two forms in the plural. All person–number combinations were distinctly marked in perfective sets, with three forms in each number (Table 6.5).[4]

While the Early Romani pattern has been retained in the majority of dialects, many dialects have extended the non-perfective second-person plural–third-person plural homonymy to some or all perfective sets (see Chapter 13

Table 6.5. Early Romani person–number suffixes

	1SG	2SG	3SG	1PL	2PL	3PL
IMPFV	*-av*	*-es*	*-el*	*-as*	*-en*	*-en*
PFV	*-om*	*-al*	*-as*	*-am*	*-an*	*-e*

Table 6.6. Latvian Romani person–number suffixes

	1SG	2SG	3SG	1PL	2PL	3PL
IMPF	*-av*	*-es*	*-el*	*-as*	*-en*	
PFV	*-um*	*-an*	*-a*	*-am*	*-e*	

for details). Thus, for example, in Latvian Romani the second-person plural and the third-person plural forms are homonymous in all finite sets; again, the corresponding singular forms are distinct (Table 6.6).

In East Ukrainian Romani (Table 6.7), and also in Podolie Romani, the category of person has been completely neutralised in the plural of perfective sets, while all three persons are kept distinct in the singular. Non-perfective sets continue the inherited Early Romani homonymy.

Table 6.7. East Ukrainian Romani person–number suffixes

	1SG	2SG	3SG	1PL	2PL	3PL
IMPFV	*-av*	*-ex*	*-el*	*-ax*	*-en*	
PFV	*-om*	*-an*	*-a*	*-e*		

Person homonymies in the singular are restricted to a few Sinti dialects. In German and Austrian Sinti, the second-person singular and the third-person singular forms are homonymous in remote sets (i.e. in the non-perfective and in the pluperfect), whereas the second-person plural–third-person plural homonymy is general (Table 6.8). Thus in some sets there are two forms in both numbers, while in other sets there are three forms in the singular and two forms in the plural. The same holds for Hungarian Sinti, except that the singular homonymy occurs in the pluperfect, but not in the imperfect.

Table 6.8. Selected German/Austrian Sinti inflections

	1SG	2SG	3SG	1PL	2PL	3PL
SUBJ	*-ap*	*-es*	*-el*	*-as*	*-en*	
IMPF	*-aw-s*	*-eh-s*		*-ah-s*	*-en-s*	
PRET	*-om*	*-al*	*-as*	*-am*	*-an*	
PLPF	*-om-s*	*-al-s*		*-am-s*	*-an-s*	

To conclude, person homonymy in singular verb inflections is always licensed by a parallel homonymy in the plural. In other words, there are no dialects with more person–number forms in the plural than in the singular.

There is also evidence for a greater person differentiation of the singular in personal pronouns. In Early Romani and numerous dialects (all Vlax and South Polish Romani), the genitive marker of the first-person singular pronoun (*-*inř*-) differs from the genitive marker of the second-person singular pronoun (*-*ir*-), while both plural pronouns have a common marker (*-*ar*-). This asymmetry between numbers has been lost in most non-Vlax dialects (see Chapter 7).

The singular also tends to be more differentiated in terms of **tense–aspect–mood** (TAM) distinctions. While in Early Romani and in most dialects all TAM sets are distinct in all person–number combinations, instances of TAM homonymy have developed in a few dialects.[5] In Hungarian Sinti there are three non-remote non-perfective sets: the subjunctive, the present, and the future (Table 6.9). They are all distinct only in the second-person singular, while other person–number combinations show some TAM homonymy. The subjunctive and the present are homonymous in the first-person singular and the third-person singular, probably due to phonological erosion.[6] In the plural, however, all three sets are homonymous due to a morphological take-over by the original future forms. Thus there are three or two forms per person–number in the singular, but only a single form in the plural.

Greater TAM differentiation of the plural is only attested in one instance. In Manuš of Auvergne the third-person singular is marked identically in the preterite and the pluperfect sets, while both sets are distinct in the other person–number combinations, including the plural ones (Table 6.10). However, the preterite–pluperfect homonymy in the third-person singular is likely to be an artefact of a surface phonotactic constraint in the dialect (cf. the expected third-person singular pluperfect *-*as-s* with a word-final geminate), and we may assume non-homonymy on the morphological level.

Table 6.9. Hungarian Sinti non-remote non-perfective inflections

	1SG	2SG	3SG	1PL	2PL	3PL
SUBJ	*-av*	*-es*	*-el*	*-ah-a*	*-en-a*	*-en-a*
PRES		*-ē*				
FUT	*-av-a*	*-eh-i*	*-el-a*			

Table 6.10. Manuš perfective inflections

	1SG	2SG	3SG	1PL	2PL	3PL
PRET	*-um*	*-al*	*-as* [*-as*]	*-am*	*-an*	*-an*
PLPF	*-um-s*	*-al-s*	*-as* [*-as-s*]	*-am-s*	*-an-s*	*-an-s*

In nouns, the plural is more likely to have more **case** distinctions than the singular. However, this asymmetry is restricted to a few inflectional classes in a few dialects. In all dialects that possess the category of case, external (Layer II) case marking shows an identical number of distinctions in both numbers. The locus of potential asymmetry is the internal (Layer I) case marking, viz. the differentiation between the nominative and the accusative (the markerless oblique case). In all oikoclitic classes, these two forms are generally distinct in both numbers. It is in xenoclitic classes that the nominative and the accusative may be homonymous in the singular; they are never homonymous in the plural. Consider the inflection of selected noun classes in East Slovak Romani (Table 6.11), where the oikoclitic classes show no homonymy, while the xenoclitic classes conflate the nominative singular and the accusative singular forms.

The nominative–accusative homonymy is typical of all Central dialects, although the individual inflections may differ. The ultimate trigger for the homonymy was the loss of distinctive stress position due to the influence of contact languages.[7] In Early Romani and many dialects, the nominative inflections of xenoclitic classes are unstressed (with the stress falling on the stem), while the accusative inflections are stressed. Frequently, the nominative singular and the accusative singular forms only differ in this suprasegmental characteristics, while segmentally they are identical. This is the case of Latvian,

Table 6.11. Case homonymy in East Slovak Romani

Class	Example	NOM.SG	ACC.SG	NOM.PL	ACC.PL
Mo	*gadž-o* ‘nonGypsy’	*-o*	*-es*	*-e*	*-en*
Fi	*bor-i* ‘daughter-in-l.’	*-i*	*-a*	*-(ij)a*	*-en*
*Mo	*sused-os* ‘neighbour’	*-os*		*-i*	*-en*
*Mi	*doxtor-is* ‘physician’	*-is*		*-a*	*-en*
*Fa	*bab-a* ‘grandmother’	*-a*		*-i*	*-en*

Finnish, and Slovene Romani for all xenoclitic classes, and in a number of Balkan dialects (e.g. Kosovo Arli, Sepečides, Sofia Erli, Yerli, Varna Bugurdži, Crimean Romani, Kosovo Bugurdži, Muzikanta, Nange) and in most South Vlax dialects (e.g. Kosovo Gurbet, Varna Kalajdži, Rešitare, and variantly Ajia Varvara) for at least some xenoclitic classes. Kosovo Bugurdži (Table 6.12) is an example of a dialect where, in the relevant singular forms, there is complete homonymy in one xenoclitic class (*MV), segmental identity but difference in stress position in two xenoclitic classes (*Mo and *Fa), and complete differentiation in another xenoclitic class (*Mi); the plural forms show no homonymy or segmental identity.

Segmental identity of the inflections, despite their different stress position, also testifies to a lesser differentiation in the singular. Finally, there are numerous dialects (e.g. Welsh Romani, Sinti, most Northeastern, North Vlax, Arli of Prilep and Florina) where the nominative and the accusative are even segmentally distinct in all inflectional classes.

Singular personal pronouns of the first and second persons tend to be more differentiated in case than the corresponding plural personal pronouns. (For case differentiation of third-person pronouns see Chapter 7). As with nouns, case homonymy only concerns internal case marking, viz. the distinction between the nominative and the accusative. In the first-person singular pronoun, the nominative *me* 'I' is always distinct from the accusative *man* (reduced *ma*). In the second-person singular pronoun, the nominative *tu* 'you' is distinct from the accusative form *tu-t*, although it coincides with the oblique stem *tu-*. In some dialects there is a reduced accusative form *tu*, which is homonymous to the nominative. The reduced accusative is a clitic variant alongside the full form in some dialects (e.g. most Vlax, Kosovo Bugurdži, and

Table 6.12. Case homonymy in Kosovo Bugurdži

Class	Example	NOM.SG	ACC.SG	NOM.PL	ACC.PL
Mo	*gadž-o* 'nonGypsy'	*-o*	*-es*	*-e*	*-en*
Fi	*bakr-i* 'sheep'	*-i*	*-ja*	*-ja*	*-jen*
*Mo	*daj-os* 'uncle'	*-os*	*-ós*	*-oja*	*-ojen*
*Mi	*oficir-i* 'officer'	*-i*	*-is*	*-ja*	*-jen*
*MV	*lovdži-s* 'hunter'	*-s*		*-da*	*-den*
*Fa	*krav-a* 'cow'	*-a*	*-á*	*-es*	*-en*

Table 6.13. Case differentiation patterns in the first-person plural pronoun

	Type A	Type B	Type C	Type D	Type E
NOM	*amen*	*amen*	*ame*	*ame*	*ame*
ACC		*amen ~ ame*	*amen*	*amen ~ ame*	

Slovene Romani), while in a few others (e.g. Venetian Sinti, Razgrad Drindari, Italian Kalderaš, or Austrian Lovari) it is the only form available. The first-person plural and second-person plural pronouns show five patterns of differentiation between the nominative and the accusative forms. Table 6.13 shows only the first-person plural pronoun, as both plural pronouns inflect alike.

Type A, with case homonymy, has been reconstructed for Early Romani (Boretzky and Igla 1994: 312, 327; Elšík 2000b: 7–8), and occurs in the Central group, many Sinti dialects, numerous Balkan dialects (e.g. Prilep Arli, Sepečides, Rumelian and Iranian Romani, Kosovo Bugurdži, Drindari, Muzikanta), and in some Vlax dialects (e.g. Ajia Varvara, Dasikano). The other types are characterised by phonological erosion of the original forms (e.g. **amen* > *ame*) in various grammatical environments. Type B is restricted to a few Balkan dialects (e.g. Arli of Gilan and Florina, or Varna Bugurdži), where the two full case forms are homonymous, while an accusative clitic is distinct from both. Type C, with no case homonymy, can be found in Welsh and Finnish Romani, some Sinti dialects, the Northeastern group, some Balkan dialects (e.g. Sofia Erli, Crimean Romani, Nange, Malokonare, Gadžikano), some South Vlax dialects (e.g. Rešitare, Kalburdžu), and Ukrainian Romani. Type D is typical of the North Vlax group (and occurs also in Yerli), where the full case forms are distinct from one another, while the accusative clitic is homonymous with the nominative. Finally, Type E, with reduced forms in all environments, shows the same pattern of case homonymy as Type A, and is attested in Varna Kalajdži.

Disregarding the clitic forms, case homonymy in the plural pronouns (Types A, B, and E) is roughly as widespread in Romani as non-homonymy (Types C and D), and it is decidedly more frequent than case homonymy in the second-person singular pronoun. Nevertheless, it is possible to find dialects where the second-person singular pronoun is less differentiated than the first-person plural and second-person plural pronouns (e.g. Austrian Lovari), so that the lesser differentiation of the plural pronouns is a matter of cross-dialectal frequency rather than implication.

Table 6.14. Inflection of xenoclitic adjectives in Early Romani

	SG.M	SG.F	PL.M	PL.F
NOM	*-o*		*-a*	
OBL	*-on-e*	*-on-a*	*-on-e*	

The singular is unambiguously more differentiated than the plural in terms of **gender** distinctions in adjectival inflection. In Early Romani, gender is neutralised in the plural of all four major inflectional classes of adjectivals (see Chapter 5). This was due to a Proto-Romani masculine take-over in the plural, which, as a sort of systematic homonymy, was retained in the new class of xenoclitic adjectives as well. In this class there is also a gender neutralisation in the singular nominative (Table 6.14), while gender forms are distinct in the singular of all the other classes.

Dialect-specific developments towards gender neutralisation in the singular are common in the oblique case, while they are extremely rare in the nominative, and so complete gender neutralisation in the singular is extremely rare, too. It is only found in Finnish Romani demonstratives, where there is no gender distinctions in the oblique, and where the original masculine nominative singular (e.g. *tauva* 'this') has extended to the feminine, becoming a gender-indifferent nominative singular form.[8] On the other hand, a few dialects have innovated the gender distinction in the plural of some adjectivals. This has occurred in the nominative of demonstratives in Hungarian Lovari (e.g. Lovari masculine *kodol-e* vs. feminine *kodol-a* 'those') and Ukrainian Romani, and with lexical adjectives in Abruzzian Romani (e.g. *-e* vs. *-ja*). In all instances the gender-differentiating inflections coincide with those of the oikoclitic vocalic noun classes (e.g. Lovari *gāž-e* 'non-Gypsy men' vs *gāž-a* 'non-Gypsy women'). Table 6.15 charts the Hungarian Lovari demonstrative inflections.

Table 6.15. Demonstrative inflection in Hungarian Lovari

	SG.M	SG.F	PL.M	PL.F
NOM	*-o*	*-i*	*-e*	*-a*
OBL	*-l-e*	*-l-a*	*-l-e*	

Table 6.16. Gender neutralisations in the third-person pronouns

	SG.M	SG.F	PL.M	PL.F
NOM (Type A)	*ov*	*oj*	*ol* (*on*)	
NOM (Type B)	*ov* (*oj*)		*on*	
OBL (Types A and B)	*l-es-*	*l-a-*	*l-en-*	

Pronouns of the third person, too, show greater gender differentiation in the singular than in the plural, due to their origin in demonstratives. Early Romani and the majority of dialects (Type A in Table 6.16) distinguish gender in the singular but not in the plural, irrespective of case. Some dialects influenced by gender-less languages (Type B) have lost the gender distinction in the nominative singular (see Chapter 8 for details). Nevertheless, this singular homonymy is still licenced by the plural one.

The only category where, in Early Romani and in some dialects, gender is encoded in verbs is the third-person singular perfective, viz. in the active participle preterite forms (e.g. *gel-o* 'he went' vs *gel-i* 'she went') and the pluperfect forms based on them (e.g. *gel-o-sas* 'he had/would have gone' vs. *gel-i-sas* 'she had/would have gone'). There is no gender distinction in the corresponding plural forms (e.g. *gel-e* 'they went' and *gel-e-sas* 'they had/would have gone').

To conclude, the greater gender differentiation of the singular may be, for all word-classes and for all dialects, formulated implicationally: if there is gender homonymy in the singular, then there is also gender homonymy in the plural.

Inflectional **classes** in verbs are more likely to be differentiated in the singular than in the plural. As for non-perfective person–number suffixes in verbs, Early Romani possessed two allomorphs in the first-person singular and in the third-person singular, while in the other categories there was a single suffix (Table 6.17).

The archaic first-person singular allomorph *-am* had been retained in ER only in a few very frequent verbs (especially *kam-am* 'I want'), and it is now

Table 6.17. Non-perfective person–number suffixes in Early Romani

1SG	2SG	3SG	1PL	2PL	3PL
-(a)v ~ -am	*-(e)s*	*-(e)l ~ -i*	*-(a)s*	*-(e)n*	*-(e)n*

Table 6.18. Inflectional class differentiation in verb inflections

	1SG	2SG	3SG	1PL	2PL	3PL
Latvian			+			
Russian, Rumelian, Malokonare, Taikon	+					
Early Romani, Gilan, Prilep, Lovari, Dasikano	+		+			
Lithuanian	+		+			+
Welsh			+			+
Polish Lovari			+	+		
Xoraxane		+	+	+		

available only in some Northeastern and some Balkan dialects (e.g. Arli of Gilan and Prilep, Rumelian Romani, and Malokonare). The Greek-derived third-person singular allomorph *-i* was, in Early Romani, used with xenoclitic verbs as opposed to *-(e)l* of oikoclitic verbs. In many dialects the xenoclitic third-person singular suffix has been completely lost, in others (e.g. Arli and Slovene Romani) it has expanded to, some or all, oikoclitic verbs (see Chapter 23). Yet in other dialects, especially Vlax, the suffix extended to other person–number categories in the inflection of xenoclitic verbs. Finally, a few dialects (e.g. Arli of Gilan and Prilep, and Xoraxane) possess a new third-person singular allomorph *-ol*. Table 6.18 summarises the presence of class differentiation for different person–number categories in individual dialects (including optional allomorphs).

In terms of number of types as well as the width of dialect distribution, the third-person singular is most likely to show class differentiation, followed by the first-person singular. Class differentiation in the plural is dialectally restricted, and licenced by class differentiation in the third-person singular. Nevertheless, the Polish Lovari and Xoraxane patterns show that, in some persons, a plural category (the first-person plural) may be more differentiated than the corresponding singular category (the first-person singular). On the whole, however, the singular non-perfective person inflections clearly show a greater tendency towards class differentiation that the plural ones.

6.4. Extension

A singular form (word, root, or affix) is more likely to extend to the plural than vice versa. There are four instances of a singular-to-plural extension (in demonstratives, verb inflection, and pronominal morphology) and a single and

controversial instance of a plural-to-singular extension (in verb inflection).

In some dialects, the nominative singular masculine form (i.e. the base form) of demonstratives is also used with non-nominative, non-singular, and/or non-masculine heads (extension to all environments turns the form into an indeclinable variant). If there are variant base forms, the extending variant is the one that is reduced in shape. The extension to the plural is always optional, i.e. specific plural forms of demonstratives are retained as well. The development has occurred independently in a number of dialects (see examples in Table 6.19).

In two dialects of the sample, the third-person singular pluperfect form or inflection extends to the corresponding plural form, the third-person plural pluperfect.[9] In Welsh Romani, the whole third-person singular form extends to the plural (e.g. *kerdīasas* 's/he has done, they had done'). There is, however, also a specific third-person plural form (e.g. *kerdenas* 'they had done'). In East Slovak Romani of Zemplín only the third-person singular inflection *-ahas*, not the whole form, extends: the third-person plural form still differs from the third-person singular form (e.g. *kerd-ahas* 'they would have done' vs. *kerď-ahas* 's/he would have done') through the lack of palatalisation of the perfective marker *-d-*. Nevertheless, the original third-person plural form in *-ehas* (e.g. *kerd-ehas* 'they would have done' < Early Romani **kerd-esas*), which is retained in closely related dialects, has been lost. In Welsh and Lithuanian Romani, the xenoclitic non-perfective third-person singular suffix *-i ~ 0* has extended to the third-person plural (see Table 6.18 in Section 6.3).

In one dialect of the sample, there has been a singular-to-plural extension in genitive pronouns. All Vlax dialects had undergone palatalisation of the second-person pronominal root **t-* in the singular genitive (Early

Table 6.19. Extension of singular demonstrative forms

	NOM.SG.M		NOM.PL	
Dialect	Full	Reduced	Specific	Extended
Welsh R	*odova*	*odā*	*odolā*	*odā*
Finnish R	*dauva*	*da*	*dāla*	*da*
Lithuanian R	*dava*	*da*	*dale*	*da*
West Slovak R	*adava*	*ada*	*ala*	*ada*
Rešitare, Kalburdžu	*ka(v)a*	–	*kala*	*ka(v)a*

Romani **t-iro* > Proto-Vlax **t'-iro* 'your.SG'), at the same time retaining the unmodified root in the plural genitive (Early Romani and Proto-Vlax **t-umaro* 'your.PL'). In Cerhari, a Vlax dialect, the palatal root has now been extended from the singular genitive to the plural genitive: *ť-umáro* 'your.PL' as well as *ť-o* 'your.SG' (Elšík 2000a:16).

According to one diachronic scenario, an extension in the opposite direction, viz. from the plural to the singular, took place in perfective inflections of the second person: the Early Romani second-person singular suffix *-al-* was replaced by the original second-person plural suffix *-an-* everywhere but in the Norwestern and the Central dialects. Since, however, the reconstruction of the second-person singular suffix *-al-* for Early Romani is controversial (cf. discussion in Matras 2002: 144), we do not know whether we can assign great importance to this counterexample.

6.5. Extracategorial distribution

Singular markers tend to have a wider distribution than categorially appropriate. Certain substantival pronouns, viz. person and thing interrogatives and indefinites (cf. Chapter 20), generally do not inflect for number.[10] Nevertheless, their oblique forms contain the singular masculine oblique marker *-s-* (*-es-*, *-as-*). Thus they are constructed as if they were singular forms. Table 6.20 shows some of the relevant pronouns.

Similarly, the Early Romani reflexive pronoun was constructed as a singular form (with the oblique stem **p-es-*, and possibly the genitive stem **p-inř-*, see Chapter 5), although it was used with plural as well as singular antecedents. This use has been retained in many dialects, including the Northeastern dialects, most North Central dialects, Slovene Romani, a number of Balkan dialects, and most South Vlax dialects. Nevertheless, numerous dialects have

Table 6.20. Singular-like oblique forms of selected pronouns

	NOM	OBL
'who'	*kon*	*k-as-* (*kon-es-*)
'what'	*so*	*so-s-*
'somebody, nobody'	*khonik*	*khanik-as-*
'something, nothing'	*či*(*či*); *khanči*	*čič-es-; khanč-es-*
'anybody, somebody'	*komoni*	*komon-es-*
'anything, something'	*čimoni*	*čimon-es-*

created specific plural reflexive forms to be used with plural antecedents (for details of their structure see Chapter 7): Welsh Romani and the Northwestern dialects, the South Central dialects, Arli of Kosovo, Sepečides, Crimean Romani, the Muzikanta, Nange and Gadžikano dialects of Sliven and Varna, and North Vlax (plus the Xoraxane dialect of Italy) with the adjacent easternmost East Slovak dialects. The use of a singular reflexive with a plural antecedent (2) and of a specific plural reflexive (3) is illustrated from Slovak Romani and the Muzikanta dialect; both forms are genitive:

(2) Slovak Romani (Lučivná)
Kada čhavoro the kaja čhajori baron avri paš ***peskri***
this little.boy and this little.girl grow.3PL out by REFL.GEN.SG:F
bibi.
aunt
'This small boy and this small girl grow up with their aunt.'

(3) Muzikanta
Kəka cikoru čəoru təj kəka cikəri čəjri barina paš ***pumari***
this little boy and this little girl grow.3PL by REFL.GEN.PL:F
bibi.
aunt
'This small boy and this small girl grow up with their aunt.'

6.6. Exposition

In the inflection of the definite article, individual category values differ with respect to the likelihood of being exposed through an individual form, not shared with other values. Disregarding regular phonological variation and dialects with an indeclinable article, the singular nominative masculine form is the most exposed (being always *o*), followed by the singular nominative feminine form (*e* or *i*), a distinct singular oblique feminine form (*ola*, *la*, or *a*), and a general oblique form (*ole*, *le*, *e*, or *i*). The nominative plural form of the definite article is the least exposed, being almost always homonymous to one of the previous forms. The interpretation of the asymmetry with regard to number is straightforward: singular forms of the article are more likely to be exposed than the plural forms. Exposition asymmetries constructed for other adjectivals do not contradict the greater exposition of the singular in the definite article.

6.7. Borrowing and internal diversity

Although borrowing of inflections (number markers as well as inflections cumulating number with other categories) has affected both numbers, borrowing of plural markers appears to be more frequent than borrowing of singular markers. This is in line with a greater internal diversity of plural inflections (especially in nouns).

In the post-Greek period, nominative plural markers in nouns have been the most likely to be borrowed (e.g. *-uri* from Rumanian, *-i* from Slavic, *-Vdes* from Greek in dialects with a prolonged contact with the language). This asymmetry is confirmed by selective borrowing of plural markers into the nominative plural of third-person pronouns in various dialects: e.g. Nógrád Rumungro *ón-k* 'they' (< Hungarian *-k*), Kalburdžu *on-lar*, Kaspičan *on-nar* and Gadžikano *on-na* (< Turkish *-lar*), or Slovene Romani *onn-i* or *jon-i* (< South Slavic *-i*). Another instance of borrowing of plural markers, viz. borrowing of the Turkic plural *-Is* into verb inflections in some dialects of the Balkans, is discussed in detail in Chapter 7.

Borrowing of singular inflections appears to be restricted to the period of the formative contact of Early Romani with Greek, and to a controversial borrowing of a first-person singular verb inflection in a single dialect (viz. Slovene Romani, see discussion in Chapter 7). The Greek-derived singular inflections borrowed into Early Romani were: the very frequent third-person singular present suffix *-i*; and the nominative inflections of what were to become xenoclitic classes in nouns (*-os*, *-is* ~ *-i*, *-a*) and adjectives (*-o*). The borrowing of the singular markers in nominals was accompanied by borrowing of corresponding plural markers.

Chapter 7
Person

The category of person is, in Romani, coded in verbs and personal pronouns. It has three values: the first person (reference to a set of people including the speaker), the second person (reference to a set of people including the addressee but not the speaker), and the third person (reference to a set of people and/or objects that does not include either the speaker or the addressee). There is no exclusive vs. inclusive distinction in Romani. In verbs, person is an agreement category that cross-references the speech-act status of the grammatical subject. On the other hand, the category is inherent in personal pronouns: the reference to speech-act participants is constitutive for their function; person constitutes their "lexical" meaning. As with number, the search for asymmetries in person is complicated by the fact that the category is rarely expressed separately. Mostly, it cumulates with other inflectional categories: with number and aspect in verbs; and with number in personal pronouns.

The first person is the most exposed, the least likely to undergo erosion, the least likely to show extracategorial distribution, and – with regard to person markers – the least prone to borrowing; it is more differentiated than the second person. The third person is, on the other hand, the most likely value to undergo erosion, show extracategorial distribution, and borrow person markers; it is ambiguous with regard to differentiation (depending on structure and cross-cutting category, it can be the most or the least differentiated value). The second person shows an intermediate tendency toward erosion, extracategorial distribution, and borrowing of person markers; it is less differentiated than the first person.

The criteria of complexity and extension render conflicting hierarchies. For any pair of the three person values, one of the values is more complex in some structures but less complex in other structures. Extensions between the first and the second persons as well as between the second and the third persons are bidirectional. Nevertheless, the mediating position of the second person is retained, in that there are no direct extensions between the first and the third persons. The criterion of borrowing renders conflicting asymmetries if borrowing of number markers into different person values is taken into account.

7.1. Complexity

The criterion of complexity reveals conflicting person asymmetries, both in verbs and in pronouns. The third person is the ambiguous value in verbs. Some developments reduce its complexity, while other developments increase its complexity. The third person can be the least as well as the most complex value, depending on the dialect and on the domain. The second person shows the least complexity in the prototypical cluster with the imperative, while otherwise the second and the first persons are on a par. In pronouns, the mutual position of the second and the first persons is ambiguous; either value can be more complex than the other, depending on the domain. We will first discuss instances of zero coding, and then other instances of structural complexity.

There are two instances of markerless forms in **verbs**. First, the category that is absolutely markerless, i.e. that corresponds to the inflectional stem, is, in Early Romani and all dialects, the second-person singular imperative of most verb classes (e.g. *ker* 'do!'). There is usually, however, no contrast with other persons: there is generally no third- or first-person singular imperative. In some dialects (e.g. East Slovak Romani), one can use a first-person plural subjunctive form in orders and suggestions directed towards a group of people that includes the speaker (e.g. *ker-as* 'we do; let's do!').

Second, some dialects have developed zero coding of the third person in the perfective past. The third-person forms are markerless relative to the corresponding first- and second-person forms, not absolutely. Moreover, the recognition of zero coding on the morphological level is partly determined by analysis. Let us first briefly summarise the relevant facts. In Early Romani and many dialects, all perfective person–number suffixes with the exception of the participial ones (i.e. third-person singular *-o* ~ *-i* and third-person plural *-e*), but including the finite third-person singular suffix *-as*, have the general shape /VC/. However, in many dialects the third-person singular suffix is now *-a* rather than *-as*. This is the case both in dialects that have undergone word-final deletion of /s/ in many other environments (e.g. the South Central dialects, Slovene Romani, Arli of Gilan and Skopje, Malokonare, Xoraxane, Priština Gurbet, Kalburdžu, and East Ukrainian Romani), and in dialects that have not (e.g. the Northeastern dialects, the easternmost varieties of Slovak Romani, Kalderaš, and Kaspičan). Considering that there is no phonological deletion of /s/ in the latter dialects, the development of **-as* to *-a* must have been morphological. Both third-person suffixes now have the general shape /V/, as against the more complex shape /VC/ of first- and second-person suffixes. We suggest that the morphological development reflects a tendency to reduce the struc-

Table 7.1. Perfective inflections in selected dialects

	1SG	1PL	2SG	2PL	3SG	3PL
Slovak R (Zemplín)	*-o-m*	*-a-m*	*-al*	*-an*	*-a*	*-e*
Polish R	*-o-m*	*-a-m*	*-a--n*	*-e--n*	*-a*	*-e*
Bunkuleš Kalderaš	*-e-m*	*-a-m*	*-a--n*	*-e--n*	*-a*	*-e*
Abruzzian R	*-o-m*	*-e--m*	*-a--n*	*-e--n*	*-a*	*-e*

tural complexity of the third-person suffix (see also below). Now, consider the perfective person–number suffixes in some of the dialects of the latter type (Table 7.1).

It is obvious that the /VC/ inflections lend themselves to further morphological segmentation. For example, the segment *-m-* occurs, in all dialects, in all first-person forms and nowhere else and may be thus considered to be a marker of the first person. Similarly, the segment *-n-* in Polish Romani, Bunkuleš Kalderaš, and Abruzzian Romani (and many more dialects) may be considered to be a marker of the second person. However, the reason why this segmentation is usually not carried out in Romani linguistics (but cf. Elšík 1997) is that the remaining segments, the vowels, do not show an absolute categorial consistency. For example in Bunkuleš, the segment *-a-* occurs in the singular of the second and the third persons but in the plural of the first person; and the segment *-e-* occurs in the plural of the second and the third persons but in the singular of the first person. Nevertheless, taking into account that some of the inflections have been affected by *morphological* extension, viz. by extension of *-e-* from the third-person plural into the second-person plural in Polish Romani and Bunkuleš, and into the second-person plural and the first-person plural in Abruzzian Romani (see Section 7.4 for details), the above segmentation gains some attractiveness. It might reflect a reanalysis of some of the suffixes into bimorphemic inflections where the consonant marks person and the vowel marks number. In Table 7.1 we have used a single hyphen to indicate consistency of person marking by the consonants, and a double hyphen to indicate that there is also consistency of number marking by

Table 7.2. Perfective inflections in Abruzzian Romani

	First person	Second person	Third person
SG	*-o-m*	*-a--n*	*-a*(*--0*)
PL	*-e--m*	*-e--n*	*-e*(*--0*)

the vowels in at least two persons. Table 7.2 gives a more suggestive layout of the inflections in Abruzzian Romani, a dialect where the bimorphemic analysis is most obvious.

The segment *-e-* now clearly marks the plural, while *-a-* is used in two out of three singular forms; *-m-* marks the first person, and *-n-* marks the second person. Importantly, the third person is markerless. This means that, on the bimorphemic analysis of the perfective person–number inflections, the third person is less complex than the first and the second persons not only in terms of the number of phonemes (viz. /V/ vs. /VC/), but also morphologically.

Another complexity asymmetry arises through the employment of the Early Romani participial suffix *-in-* (see Chapter 21 for details). In the copula, third-person forms are more likely to contain the suffix than first- and second-person forms. There are dialects where the suffix is restricted to the third person (e.g. East Slovak Romani *h-in* 's/he is' but *s-om* 'I am', Karditsa Arli (*i*)*s-in-es* 's/he was' but (*i*)*s-om-as* 'I was'), but no dialects where it is restricted to the first and second persons. On the other hand, in a few dialects the suffix occurs in all forms of the first and second persons, but only in the past of the third person (e.g. Erli *s-in-jom* 'I am', *s-in-jom-as* 'I was', *s-in-e* 's/he was', but *s-i* 's/he is'; see also Chapter 13). Similarly, in perfective verbs, the suffix *-in-* is more likely to occur in the adjectival form of the third-person plural than in the other forms (e.g. Lovari *d-in-e* 'they gave' but *d-em* 'I gave').

In **personal pronouns**, there is a tendency for inflectional markers of the first person to be more complex (morphologically or phonologically) than those of the second person. This is only the case in the singular pronouns, while the plural pronouns show a completely parallel inflection. On the other hand, a couple of dialects have developed zero root marking of the first person with regard to the second person in plural pronouns (see below).

There are three instances of a lesser complexity of the first-person singular inflections. First, the oblique stem of the first-person singular pronoun is, in all dialects, derived by means of the suffix *-an-*, while there is no overt oblique marker in the second-person singular pronoun (cf. *m-an-* vs. *tu-*). Second, in Early Romani, the genitive (possessive) stems of the first-person singular and the second-person singular pronouns were **m-inř-* and **t-ir-*, respectively, and so the first-person singular genitive suffix **-inř-* was longer (i.e. phonologically more complex) than the second-person singular genitive suffix **-ir-*. This pattern has been retained in some Vlax dialects (e.g. Varna Kalajdži *m-əndr-* vs. *k-ir-*, Ukrainian *m-ern-* vs. *t'-ir-*) and in the North Central dialects of southern Poland and northeastern Slovakia (*m-indr-* vs. *t-ir-*). In other Vlax dialects the genitive suffixes are still different in both pronouns, but none of

them is demonstrably more complex than the other (e.g. Rešitare *m-un-* vs. *k-ir-*, Dasikano *m-rn-* vs. *ć-ir-*). Most non-Vlax dialects have undergone a morphological unification of genitive marking in the singular pronouns (see Section 7.4), and so they show no difference in complexity between the first and the second persons (e.g. Sepečides *m-indr-* and *t-indr-*). A third instance of a greater complexity of the first-person singular pronoun concerns reduced variants of singular genitives (possessives), which have developed in many dialects, including Lovari. In Lovari, the reduced first-person singular variant *m-* is less frequent than the corresponding unreduced variant *m-ur-*, while the reduced second-person singular variant *t-* is more frequent than the corresponding unreduced variant *t-ir-*. In some Lovari varieties the more frequent variants won the competition, and the resulting opposition between the first-person singular *m-ur-* and the second-person singular *t-* again shows a greater complexity of the first-person genitive.

In some Northwestern dialects (e.g. in Austrian, Hungarian and Lombardian Sinti, and in some varieties of Finnish Romani), all forms of the first-person plural pronoun have been affected by phonological erosion of unstressed initial vowels: **ame(n)*, oblique **amen-*, genitive **amar-* > *me(n)*, *men-*, *mar-*. What has been lost was the actual root *a-* of the first-person plural pronoun. The first-person plural forms now possess a zero root with regard to the second-person plural forms: cf. the inflectional stems first-person plural *0-m-* vs. second-person plural *tu-m-* (where *-m-* is an irregular plural marker). The phonological development hence resulted in morphological zero marking, in the plural pronouns, of the first person with respect to the second person.

Another development that can be discussed under the criterion of complexity is the alternation of the roots *s-* and *h-* in indicative copula forms.[1] Some dialects possess only a single root in all grammatical environments. Thus Welsh Romani, the Northeastern dialects, Abruzzian Romani, most Balkan dialects, and most Vlax dialects, including Ukrainian Romani, always employ the root *s-* (e.g. Sofia Erli *s-i* 's/he is, they are', *s-ine* 's/he was, they were', and *s-injom* 'I am'); and the Core Sinti dialects, Slovene Romani, and Cerhari always employ the root *h-* (e.g. Slovene Romani *h-i* 's/he is, they are', *h-ine* 's/he was, they were', and *h-injum* 'I am'). Both roots are used variantly in all inflectional environments in the Arli dialects of Gilan and Skopje, and in Piedmontese Sinti (e.g. Skopje Arli *s-i* ~ *h-i* 's/he is, they are', *s-ine* ~ *h-ine* 's/he was, they were', and *s-injum* ~ *h-injum* 'I am'). However, in some dialects, both roots are employed in such a way that their distribution is determined by values of inflectional categories. Concretely, there may be a distinction between the roots in the third person of the present, in the third person of the

past, and in the first and the second persons of either tense. Table 7.3 shows three types of distribution of the two roots.

In Type A, the root *s-* is unrestricted in its distribution, but there is also a variant in *h-* in the third-person present. This type is found in the South Central dialects, Prizren, Karditsa, and Xoraxane (e.g. Prizren *s-i ~ h-i* 's/he is, they are' vs. *s-ine* 's/he was, they were' and *s-im* 'I am'). Type B, where the root *h-* occurs only in the third-person present, and the root *s-* elsewhere, is attested in Finnish Romani, the easternmost varieties of Slovak Romani, and Dasikano (e.g. Finnish Romani *h-in* 's/he is, they are' vs. *s-as* 's/he was, they were' and *s-om* 'I am'). Finally, in Type C, the root *h-* occurs in third-person forms of either tense, while the root *s-* occurs in the other persons. This pattern is found in most North Central dialects, from Bohemia in the west to the Zips region of East Slovakia in the east (e.g. Zips Slovak Romani *h-in* 's/he is, they are' and *h-as* 's/he was, they were' vs. *s-om* 'I am').

There is a clear implicational pattern in the distribution of the roots. Assuming the perspective of the root *h-*, for example: its presence in the first- and second-person forms implies its presence in the third-person forms; and its presence in the past third-person form implies its presence in the present third-person form. As for person asymmetry, there is a clear split between the third person on the one hand, and the first and the second persons on the other hand. Now, the synchronic evaluation of this asymmetry depends on historical interpretation of the developments. It is possible that both roots were inherited into Early Romani as variants, and that there was, in different dialects, either a generalisation of one root in all environments, or else a re-distribution of the two roots according to person and/or tense (proposing this scenario, Matras 1999c termed it *option selection*). Or it is possible that there was a single root *s-*, which, in some dialects, in some or all grammatical environments, eroded to *h-*. If we assume the erosion scenario, then we obtain the following person hierarchy of erosion: 3 > 1, 2. If we evaluate the option selection scenario, then we obtain an opposite hierarchy of phonological complexity (assuming a greater complexity of the 'strong' root *s-* with regard to the 'weak' root *h-*): 1, 2 > 3.

Table 7.3. Categorially determined distribution of indicative copula roots

	Third-person present	Third-person past	First and second person
Type A	*s- ~ h-*	*s-*	*s-*
Type B	*h-*	*s-*	*s-*
Type C	*h-*	*h-*	*s-*

Both hierarchies, or rather the criteria that underlie them, are clearly complementary: the more eroded a form is, the less complex it is, and vice versa.

7.2. Erosion

The person hierarchy with regard to phonological erosion is 3 > 2 > 1. We discuss two developments where the third person is most likely to undergo erosion: contractions and other erosion developments in middle verbs, and reduction of the remoteness suffix. The relative position of the second and the first persons on the above hierarchy is only obvious from the patterns of erosion in middle verbs.

In most dialects, erosion has affected the non-perfective inflection of middle verbs in *-(j)ov-*. Inflections of different person–number values were affected to a differing degree in different dialects. In Early Romani, the middle verbs inflected like other consonantal stems: they contained the middle suffix **-jov-* and the regular person–number suffixes of the shape /VC/ (e.g. subjunctive **bar-jov-av* 'I grow', **bar-jov-es* 'you.SG grow', **bar-jov-el* 's/he grows'). The locus of erosion has been the sequence of the middle suffix and the following vowel of the person–number suffixes (i.e. **-jov-a-* in the first person, and **-jov-e-* in the second and third persons). Since the initial yod of the middle suffix has frequently fused with preceding consonants or has been deleted altogether (e.g. West Slovak Romani *šund'-ov-* < **šund-jov-* 'be heard', or *bar-ov-* < **bar-jov-* 'grow'), we will leave it out of consideration for the most part. For convenience, we will term the sequence of the **-ov-* of the middle suffix and the following vowel of the person–number suffixes as the *middle sequence*. Table 7.4 shows reflexes of the Early Romani middle sequences **-ova-* and **-ove-*, classified according to the number of phonemes they contain (which will enable us to evaluate the degree of erosion).

The relevant processes of erosion of the middle sequences have been, for example: vowel raising (e.g. **-ove-* > *-uve-*), deletion of the intervocalic /v/

Table 7.4. Erosion of the middle sequences

Person	Early Romani	3 phonemes	2 phonemes	1 phoneme
First	**-ova-*	*-uva-, -iva-, -oja-*	*-oa-, -ia-*	*-a-*
Second/third	**-ove-*	*-uve-, -ive-*	*-oe-, -oj-*	*-o-, -u-, -i-*

(e.g. **-ove-* > *-oe-*), and consequent contraction of the vowel sequence (e.g. *-oe-* > *-o-*). The yod of the middle suffix has participated in the development of the middle sequences beginning in /i/ (e.g. **-jove-* > *-ive-* > **-ie-* > *-i-*). There are at least 25 distinct combinations of different middle sequences in various dialects. Table 7.5 shows 14 more abstract erosion patterns, based on the number of phonemes in the middle sequence (indicated by digits). Two sorts of shading are used to visualise differing degrees of erosion.

No erosion of the middle sequences (Type A) has occurred in Slovene Romani and older Finnish Romani (e.g. Slovene Romani *barj-ova-m* 'we grow', *barj-ove* 'you.SG grow', *barj-ovi* 's/he grows'). In modern Helsinki Romani (Type B), contraction optionally affects the third-person singular form (e.g. *parj-uve-la* ~ *parj-u-la* 's/he grows'). Piedmontese Sinti is a dialect of Type C: third-person forms of both numbers are obligatorily contracted. In the dialect of Vălči Dol (Type D), the third-person singular form is always contracted (e.g. *barj-o-l* 's/he grows'), while there is variation in the second/third-person

Table 7.5. Patterns of erosion in the middle sequences

	1PL	1SG	2SG	2PL	3PL	3SG
Type A	3	3	3	3	3	3
Type B	3	3	3	3	3	3/1
Type C	3	3	3	3	1	1
Type D	3	3	3	3/1	3/1	1
Type E	3	3	3	1	1	1
Type F	3	3	2	1	1	1
Type G	3	3	3/1	3/1	1	1
Type H1	3	3	3/1	3/1	3/1	3/1
Type H2	3	3	1	1	1	1
Type I	3 (ova)	3 (iva)	1	1	1	1
Type J1	2	2	2/1	2/1	2/1	2/1
Type J2	2	2	1	1	1	1
Type J3	3/1	3/1	1	1	1	1
Type K	1	1	1	1	1	1

plural (e.g. *bar-ive-n* ~ *barj-o-n* 'you.PL/they grow'). The obligatory contraction in the third-person singular and second/third-person plural forms (Type E) is attested in West Slovak Romani and some modern varieties of Finnish Romani. Roman (Type F) shows erosion in all forms: in the first-person forms, however, the middle sequence retains three phonemes (e.g. *barč-oja-v* 'I grow'), while it consists of two phonemes in the second-person singular (e.g. *barč-oj-s* 'you.SG grow'), and a single phoneme in the other forms (e.g. *barč-o-l* 's/he grows'). Šóka and Klenovec Rumungro and Prizren Arli (Type G) exhibit obligatory contractions in the third person, and optional contractions in the second person (e.g. Šóka *bārď-o-n* 'they grow' vs. *bārď-ove-n* ~ *bārď-o-n* 'you.PL grow'). In Type H, forms of the first person show no erosion, while the other forms are contracted. An optional contraction (subtype H1) is found in Arli of Skopje and Gilan, Sofia Erli, Rumelian Romani, Taikon Kalderaš, and Xoraxane; an obligatory contraction (subtype H2) is attested in the Northeastern dialects, Bohemian and East Slovak Romani, Markuleš Kalderaš, Lovari, Dasikano, and Varna Kalajdži (e.g. Latvian Romani *bārj-uva-v* 'I grow' vs. *bārj-u-s* 'you.sg grow' and *bārj-u-l* 's/he grows'). In Kalburdžu (Type I), only the middle sequence in the first-person plural form does not show any erosion (e.g. *barj-ova* 'we grow'); the first-person singular middle sequence retains three phonemes but the yod of the middle marker has been deleted (e.g. *bar-iva-v* 'I grow'), and there is full contraction in the other persons (e.g. *barj-o* 'you.SG grow'). All forms are affected by erosion in Type J; however, the first-person forms are less eroded than the forms of the other persons. In Florina Arli (subtype J1) and Ajia Varvara (subtype J2), there are two segments in the first-person middle sequence (e.g. Florina *barj-oa-va*, Ajia Varvara *bar-ia-v* 'I grow'), and an optional or obligatory full contraction in the other persons (e.g. Florina *barj-o*(*e*)*-sa*, Ajia Varvara *bar-o-s* 'you.SG grow'). In Prilep Arli (subtype J3), there is an optional contraction in the first person, and an obligatory contraction in the other persons. Finally, a number of dialects exhibit Type K, viz. full contraction in all forms (e.g. Crimean Romani *bar'-a-v* 'I grow', *bar'-o-s* 'you.SG grow', *bar'-o-l* 's/he grows'). This type is found in Welsh Romani (in a few verbs with petrified middle morphology), in a great number of Balkan dialects (e.g. Sepečides, Yerli, Varna Bugurdži, Crimean Romani, Kosovo Bugurdži, Varna Gadžikano, Kaspičan, Malokonare, Nange, Muzikanta), and in some Vlax dialects (e.g. Bunkuleš Kalderaš and Rešitare).

The patterns of erosion shown in Table 7.5 suggest the following hierarchy of erosion: third-person singular > third-person plural > second-person plural > second-person singular > first-person singular > first-person plural. The hierarchy is implicational: if a value shows some erosion in the middle

sequence, then all the values to the left on the hierarchy will show the same or a greater degree of erosion. As far as the category of person is concerned, the interpretation of the hierachy is straightforward: third-person forms tend to undergo more erosion than second-person forms, which in turn tend to undergo more erosion than first-person forms (3 > 2 > 1). However, it must be noted that the Early Romani first-person middle sequence (*-*ova*-) is phonologically 'stronger' than the second- and third-person sequence (*-*ove*-), and so the tendency of first-person forms to be least affected by erosion might be partly attributed to their more 'favourable' starting position.[2]

The Early Romani remoteness suffix *-*asi* has been eroded in all dialects (to -*as* in most dialects, and further to -*es* or -*s* in Core Sinti, -*ah* in Dasikano, and -*a* in Xoraxane; and to -*ahi* in South Central dialects and further to -*ai* in Prizren). The erosion has mostly had the same outcome in all grammatical environments. In some Northeastern dialects, however, we find further erosion of the suffix (*-*as* > -*ys* in Lithuanian and Russian Romani, and further to -*is* in Latvian and Estonian Romani) only in certain environments. In Estonian, Latvian, Lithuanian, and Russian Romani, we find (*i*)*s-ys* or (*i*)*s-is* < **is-as* 's/he was, they were' in the third person of past copula forms, but only the unreduced (*i*)*som-as* 'I was', (*i*)*san-as* 'you were' etc. in the other persons. In Latvian, Lithuanian, and Russian Romani, but usually not in Estonian Romani, the reduced variant of the remoteness suffix has further spread to the third-person pluperfect forms of lexical verbs, which show significant structural parallelism to the third-person past forms of the copula (e.g. *kerd'a-s-ys* < **kerdjas-as* 's/he had done' and *kerde-s-ys* < **kerde-s-as* 'they had done'). The unreduced variant of the suffix has been retained in the other persons of the pluperfect (e.g. *kerd'om-as* 'I had done', *kerd'an-as* 'you had done' etc.), as well as in imperfect forms of all persons (e.g. *kerel-as* 's/he was doing'). Unlike the above mentioned dialects, Polish Romani has eroded the remoteness suffix in all grammatical environments, irrespective of person (i.e. also in

Table 7.6. Variants of the remoteness suffix in the Northeastern dialects

Dialect	3RD PAST (cop)	3RD PLPF (verbs)	Elsewhere
Estonian R	-*is*	-*as*	-*as*
Latvian R	-*is*	-*is*	-*as*
Lithuanian–Russian R	-*ys*	-*ys*	-*as*
Polish R	-*ys*	-*ys*	-*ys*

som-ys 'I was', *kerdžom-ys* 'I had done', *kerel-ys* 's/he was doing' etc.). Table 7.6 summarises the distribution of the remoteness variants in the Northeastern dialects; cells with reduced variants are shaded.

It is obvious from Table 7.6 that the third person is more likely to be affected by erosion than the other persons. The person hierarchy here is 3 > 1, 2.

7.3. Differentiation

Various person asymmetries emerge when employing the criterion of differentiation, depending on which cross-cutting category is considered, and also on what sort of diachronic process has been involved. The third person is clearly the most differentiated value in terms of case, gender, and class distinctions. It is also the most differentiated value in terms of TAM distinctions, whenever they have a morphosyntactic origin. On the other hand, loss of TAM distinctions due to phonological erosion affects the third person most. As for differentiation in number, the third person is the least differentiated in personal pronouns, and it shows medium differentiation in verbs. The first person appears to be the most differentiated value in terms of number distinctions and irregularity of coding. In verbs, interestingly, the second person shows the fewest number distinctions. However, various dialect-specific developments tend to 'improve' the position of the second person in this respect. The first and the second persons frequently do not exhibit any mutual ranking. This is so with regard to case and gender distinctions, and with regard to TAM distinctions of morphosyntactic origin. Table 7.7 summarises the various person asymmetries.

Below we discuss person differentiation in individual cross-cutting categories: number, TAM, case, and gender. Class differentiation is only relevant for verbs.

In verb inflection, the first person is most likely to contain **number** distinctions, while in the third person and especially in the second person number can sometimes become neutralised. Number neutralisation in the first person is

Table 7.7. Differentiation asymmetries in the category of person

	Number	TAM	Case	Gender	Class
Verbs	1 > 3 > 2	3 > 1, 2 (1 > 2 > 3)	–	3 > 1, 2	3 > 1 > 2
Pronouns	1 (>) 2 > 3	–	3 > 1, 2	3 > 1, 2	–

restricted to current varieties of Finnish Romani.[3] Although number neutralisation is most common in the second person, there is a tendency toward secondary differentiation of number through various developments. All number homonymies are restricted to perfective sets (see Chapter 13).

The Early Romani distinction between the perfective second-person singular *-al* and second-person plural *-an* has been retained in older Finnish Romani, the Sinti dialects, and the Central dialects. The Sofia Erli paradigm (Table 7.8) represents dialects that do not differentiate number in the second person of perfective sets (e.g. *kerdj-an* 'you.SG/PL did').

The same pattern (with the second person suffix *-an-*) is also found in Polish and Hungarian Lovari, Taikon Kalderaš, and Vălči Dol. In Austrian Lovari, number is neutralised in the pluperfect (e.g. *kerd-an-as* 'you.SG/PL had/would have done'), and only variantly in the preterite: there is a form that may be used in both numbers (e.g. *kerd-an* 'you.SG/PL did'), but also a specifically singular form (e.g. *kerd-al* 'you.SG did').

A number of dialects that originally must have shown the pattern given above for Sofia Erli have applied various means to secondarily distinguish number in the second person. The most common development to this end has been the change of the second person suffix **-an-* > *-en-* in the plural (i.e. second-person singular *-an-* vs. second-person plural *-en-*), due to an influence of the third-person plural suffix (see Section 7.4).[4] This has occurred in Polish and Abruzzian Romani, in the majority of Balkan dialects (e.g. in Arli of Gilan, Prilep, and Florina, in Sepečides, Varna and Kosovo Bugurdži, Iranian Romani, Gadžikano, Muzikanta), and in most Vlax dialects of the southern Balkans (e.g. in Bunkuleš Kalderaš, Dasikano, Priština Gurbet, Ajia Varvara, Varna Kalajdži, Rešitare). The development is underway in Xoraxane: the old number-indifferent form (e.g. *čerd-an* 'you.SG/PL did') is supplemented by a new, specifically plural, form (e.g. *čerd-en* 'you.PL did'). A second way to distinguish the second-person forms has been to suffix a Turkic-derived number marker onto the second-person plural form (e.g. Crimean Romani *kerd'-an* 'you.SG did' vs. *kerd'-an-us* 'you.PL did'). This development will be discussed in detail in Section 7.7.

Table 7.8. Sofia Erli perfective inflections

	1SG	1PL	2SG	2PL	3SG	3PL
PRET	*-om*	*-am*	*-an*		*-as*	*-e*
PLPF	*-om-as*	*-am-as*	*-an-as*		*-as-as*	*-e-s-as*

Table 7.9. Rumelian Romani perfective inflections

	1SG	1PL	2SG	2PL	3SG	3PL
PRET	*-om*	*-am*	*-an*		*-a(s)*	
PLPF	*-om-as*	*-am-as*	*-an-as*			

In Rumelian Romani, both the second and the third persons show number neutralisation in the perfective sets (Table 7.9): the number homonymy in the second person described above for Sofia Erli, and an additional number homonymy in the third person due to an extension of the third-person singular marker into the third-person plural (see Chapter 6).

`The perfective paradigm of Welsh Romani is similar to that of Rumelian Romani, except that the third-person homonymy is optional and restricted to the pluperfect set: the preterite forms are distinct (e.g. *kerd-as* 's/he did' vs. *kerd-e* 'they did'), and alongside the number-indifferent pluperfect form (e.g. *kerd-as-as* 's/he/they had/would have done') there is also a specifically plural form (e.g. *kerd-en-as* 'they had/would have done'). The paradigms of Rumelian and Welsh Romani appear to suggest that number homonymy in the third person is licenced by number homonymy in the second person. However, we also find dialects with number homonymy in the third person alone. Consider the perfective paradigm of Slovak Romani of the Zemplín region:

In this dialect, the third-person singular inflection has extended into the third-person plural in the pluperfect set, whereas the second-person inflections are distinct in all sets. However, the actual third-person pluperfect forms do differentiate number through palatalisation of the perfective marker (e.g. *kerď-ahas* 's/he would have done' vs. *kerd-ahas* 'they would have done'; see also Chapter 6), and so there is only homonymy of inflections, not of the whole word-forms. A dialect with a genuine number homonymy in the third person without a corresponding homonymy in the second person is Bougešťi Lovari (see Table 7.11).

Table 7.10. Slovak Romani (Zemplín) perfective inflections

	1SG	1PL	2SG	2PL	3SG	3PL
PRET	*-om*	*-am*	*-al*	*-an*	*-a*	*-e*
PLPF	*-om-as*	*-am-as*	*-al-as*	*-an-as*	*-a-h-as*	

Table 7.11. Bougešťi perfective inflections

	1SG	1PL	2SG	2PL	3SG	3PL
PRET	*-em*	*-am*	*-al*	*-an*	*-a*	*-e*
PLPF	*-em-as*	*-am-as*	*-al-as*	*-an-as*	*-a-s-as*	*-e-s-as*
IRR	*-em-as*	*-am-as*	*-al-as*	*-an-as*	*-oun*	

In the third person, the dialect differentiates two subsets of the pluperfect: the first one conveys a pluperfect meaning (e.g. *kerd-esas* 'they had done'), while the second one is used as an unreal conditional or irrealis (e.g. *kerd-oun* 's/he/they would have done'). In the irrealis subset, the third person does not differentiate number. The second person, on the other hand, does differentiate number, although there is no distinctions of the two pluperfect subsets.[5] Thus, it can be concluded, number homonymy in the third person alone either only concerns inflections, not word-forms (in Zemplín Romani), or it is restricted to very specific TAM contexts (in Bougešťi). On the other hand, number homonymy in the second person alone does not show any such restrictions, and is moreover much more widespread cross-dialectally.

Varieties of Finnish Romani, show significant variation in their person–number suffixes in the perfective. Table 7.12 shows the patterns attested in our sample:

In earlier Finnish Romani (Type A), which retained the Early Romani inflections, there was no number homonymy at all. Varieties of Types B, C, and D have extended the original third-person plural suffix *-e* to the second-person plural as well, which did not affect the distinction in number in the second person. This has been the only extension in varieties of Type B, and so there is no number homonymy either. In varieties of Types C and D, however,

Table 7.12. Finnish Romani preterite inflections

	1SG	1PL	2SG	2PL	3SG	3PL
Type A	*-om*	*-am*	*-al*	*-an*	*-as*	*-e*
Type B	*-om*	*-am*	*-al*	*-e*	*-as*	*-e*
Type C	*-om*		*-al*	*-e*	*-as*	*-e*
Type D	*-om*		*-al*	*-e*	*-as*	
Type E	*-om*	*-omm-as*	*-al*	*-all-as*	*-as*	*-e*

there has also been an extension of the first-person singular suffix *-om* into the first-person plural, resulting in number neutralisation in the first person (e.g. *cert-om* 'I/we did'). In addition, the third-person singular suffix *-as* extends into the third-person plural in varieties of Type D, which brings about number neutralisation in the third person. Finally, in varieties of Type E, there was an extension of the second-person singular suffix *-al* to the second-person plural, as well as the extension of the first-person singular suffix *-om* to the first-person plural (as in Types C and D). However, the resulting number homonymy in the first and the second persons has been removed by an exaptation of the remoteness suffix *-as* to mark the plural (e.g. *cert-omm-as* 'we did' < *'I had/would have done'). Consequently, the form without the remoteness suffix has been assigned the singular function (e.g. *cert-om* 'I did').[6] Type C engenders the differentiation person hierarchy 2, 3 > 1; Type D engenders 2 > 1, 3; and Type E engenders 3 > 1, 2. The generalisations that can be gained on the basis of Finnish Romani data is that the second person and the third person tend to be more differentiated than the first person, which is exactly the opposite of what emerges from the data in other Romani dialects. We suggest that the 'misbehaviour' of current Finnish Romani might be attributable to language obsolescence.

Pronouns of different persons show asymmetry in terms of their differentiation for number. Although there is no complete number neutralisation in any person, pronouns differ with regard to irregularity of number marking. In Early Romani and most dialects, first-person pronouns exhibit strong number suppletion of their roots (cf. first-person singular *m-* vs. first-person plural *a-*), while the roots of second and third-person pronouns are not suppletive across number: cf. the second-person root *t-*, and the third-person roots *o-* (nominative) ~ (*o*)*l-* (oblique). Unlike the third-person pronouns, the second-person pronouns show some irregularity in their stem formation: the second-person plural pronoun contains a stem formative *-u-* in all of its forms (and thus it is part of its base stem), while in the second-person singular pronoun this formative does not occur in the genitive stem **t-ir-* (and thus it is not part of its base stem). Furthermore, first and second-person pronouns employ an irregular plural marker *-m-*, while in third-person pronouns number marking is irregular in the nominative, but completely regular in the oblique forms. Thus, in Early Romani and most dialects, the person hierarchy of irregularity (and hence differentiation) in number marking is 1 > 2 > 3.

In some Northwestern dialects, all forms of the first-person plural pronoun have been affected by phonological erosion of unstressed initial vowels (see Section 7.1). In some of these dialects, the first-person plural form is now,

at least variantly, homonymous with the first-person singular form (cf. first-person singular *me* and first-person plural *me* < **ame*). Number remains distinguished in the other cases (cf. oblique first-person singular *man-* vs. first-person plural *men-* < **amen-*). Although the nominative forms of the first-person pronouns now show number homonymy on the surface, one could argue that they are still distinct on a morphological analysis: the segment *m-* is a root cumulating person and number in the first-person singular pronoun, while it is a plural marker in the first-person plural pronoun. On the other hand, it cannot be excluded that reanalysis has taken place, whereby the segment *m-* in the first-person singular and the first-person plural pronouns has been reinterpreted as a first-person marker. Number differentiation would be then carried by the vowels of the oblique inflection (cf. first-person singular *m-a-n-* vs. first-person plural *m-e-n-*). Thus, in these dialects, the position of the first person on the hierarchy of number irregularity appears to co-vary with a particular morphological analysis of the first-person forms.

There are conflicting asymmetries with regard to differentiation of different persons in **TAM** categories. In some instances, the third person is clearly more differentiated than the other persons, while in other instances, it is the least differentiated. We will first consider instances of a greater differentiation of the third person.

First, as described above, Bougešťi and Austrian Lovari differentiate two subsets of the pluperfect: a genuine pluperfect and an unreal conditional (irrealis). An overt distinction between these two subsets is restricted to the third person (e.g. *kerd-asas* 's/he had done' vs. *kerd-oun* 's/he/they would have done'), while in the other persons the pluperfect and the irrealis are homonymous (e.g. *kerd-omas* 'I had/would have done').

Second, certain verbs in some dialects show form differentiation in the third-person singular of (some or all) perfective sets: beside the gender-indifferent finite forms in **-as-*, there are also the so-called active participle forms that encode gender of the subject (e.g. Lovari *gelo* 'he went' and *geli* 'she went' beside *gelas* 's/he went'). Some dialects (e.g. Welsh Romani, the Northeastern dialects, Sinti, and most North Central and Rumungro varieties) have generalised the finite form, and so there is no differentiation in the third-person singular. In some dialects each verb allows only one of the third-person singular forms (e.g. in Kosovo Bugurdži verbs of movement, state, or change of state use the active participle form, while other verbs employ the finite form), and so there is differentiation in verb classes, but no intraparadigmatic differentiation in the inflection of individual verbs. Dialects that allow both options with at least some verbs, and which are thus of relevance here, include some

Table 7.13. Perfective inflections in earlier Finnish Romani

	1SG	2SG	3SG	1PL	2PL	3PL
PRET	*-om*	*-al*	*-as*	*-am*	*-an*	*-e*
PLPF	*-om-as*	*-al-as*		*-am-as*	*-an-as*	

Balkan dialects (e.g. Arli of Gilan, Skopje, and Prilep, and Sepečides), North Vlax, some South Vlax dialects (e.g. Priština Gurbet, Xoraxane, Ajia Varvara, and Vălči Dol), the Vendic dialects, earlier Rumungro varieties, and some varieties of Finnish Romani. At least in some of these dialects, the distinction between the finite and the active participle forms encodes evidentiality (cf. Matras 1994a for Lovari/Kelderaš). Whatever the functional differentiation between the finite forms and the active participles, it is significant that the formal differentiation is restricted to the singular of the third person. The other persons possess no differentiation of this sort.

In some other instances the third person is the least differentiated of all persons. First, in earlier Finnish Romani, the third-person preterite suffixes have extended into the pluperfect as well, which resulted in a preterite–pluperfect homonymy in the third person (Table 7.13).

Similar patterns are attested in Rumelian Romani, and in the third-person singular also in Manuš, where however we have argued for morphological non-homonymy (see Chapter 6). Second, in some Sinti dialects, certain TAM distinctions in non-perfective sets have been neutralised due to phonological erosion. Consider, for example, the Austrian Sinti paradigm (Table 7.14).

Historically, the Sinti present-future1 forms arose through a split phonological development of the original present–future set in *-a*: the suffix survived in some syntactic environments (the current present–future2 set), while it has been lost through erosion in others (the current present-future1 set); see also Chapter 13. In the third-person singular and the second/third-person plural, the erosion resulted in homonymy with corresponding subjunctive inflections:

Table 7.14. Non-remote non-perfective inflections in Austrian Sinti

	1SG	2SG	3SG	1PL	2PL	3PL
SUBJ	*-ap*	*-es*	*-el*	*-as*	*-en*	
PRES-FUT1	*-aw*	*-eh*		*-ah*		
PRES-FUT2	*-av-a*	*-eh-e*	*-el-a*	*-ah-a*	*-en-a*	

Table 7.15. Non-remote non-perfective inflections in Hameln Sinti

	1SG	2SG	3SG	1PL	2PL	3PL
SUBJ	*-ap*	*-es*	*-el*	*-as*	*-en*	
PRES-FUT1	*-aw*	*-eh*		*-ah*		
PRES-FUT2	*-aw-a*		*-el-a*		*-en-a*	

e.g. third-person singular subjunctive *-el* (< **-el*) as well as present-future1 *-el* (< **-el-a*). In other person–number combinations, the distinction has usually been maintained: e.g. first-person singular subjunctive *-ap* (< **-av*) vs. present-future1 *-aw* (< **-av-a*). The first person remains differentiated in both numbers and the second person at least in the singular, while in the third person there is now a TAM homonymy in both numbers. The pattern is more complex in current German Sinti, however (Table 7.15).

Here there is both the homonymy between the subjunctive and the present-future1 sets (as described above), and a homonymy between both present-future sets. In all likelihood, the latter homonymy arose through erosion of the marker *-a* in the present-future2 set rather than through a morphological extension. Each homonymy affects different person–number combinations, and the only generalisation that may be formulated is that, in the singular, the first person tends to retain most distinctions. This is also confirmed by Manuš, where the first-person singular is the most differentiated in terms of TAM distinctions (Table 7.16).

Like the other Sinti dialects above, Manuš shows homonymy between the subjunctive and the present-future1 sets; it does not show the homonymy between both present-future sets (as found in German Sinti). However, there is a homonymy between the subjunctive and the imperfect forms of the second-person singular and first-person plural, also due to erosion: e.g. first-

Table 7.16. Non-perfective inflections in Manuš

	1SG	2SG	3SG	1PL	2PL	3PL
PRES-FUT1	*-aw* (*-ō*)	*-e*	*-el*	*-a*	*-en*	
SUBJ	*-ap*	*-es* [-es]		*-as* [-as]		
IMPF	*-o-s*	*-es* [-e-s]	*-el-s*	*-as* [-a-s]	*-en-s*	
PRES-FUT2	*-ov-a*	*-er-a*	*-el-a*	*-ar-a*	*-en-a*	

person plural subjunctive *-as* (< **as*) vs. imperfect *-a-s* (< **-ah-s* < **-ah-as* < **-as-as*). Nevertheless, one can argue for a morphological distinction between the phonologically homonymous forms (see the segmentation in Table 7.16). The generalisation over the Sinti data in Tables 7.14–16 seems to be that the third person is the most likely, and the first person the least likely, to develop TAM homonymies through erosion. Hence, the asymmetry with regard to TAM differentiation is: 1 > 2 > 3.

The extent of **case** differentiation in personal pronouns reveals a clear person asymmetry: the third person is more differentiated than the first and the second persons. The examination of case distinctions in the first and second-person pronouns in Chapter 6 (Section 6.3) may be summarised as follows: the first-person singular pronoun never shows case homonymy; the second-person singular pronoun shows case homonymy only in a few dialects; and case homonymy is common in the first-person plural and second-person plural pronouns. This asymmetry in case differentiation (first-person singular > second-person singular > first-person plural, second-person plural) cannot be translated neatly into a person hierarchy. The greater differentiation of the first person with regard to the second person is restricted to the singular. The third-person pronouns not only do not have any case homonymy, but their case forms show irregularity, and even root suppletion in some dialects. Consider the paradigm of third-person pronouns in selected Balkan dialects (Table 7.17).

The Crimean Romani paradigm represents the situation in most (viz. all non-Balkan and some Balkan) dialects: the nominative root *o-* (or *jo-* or *vo-*, according to dialect) contrasts with the oblique root *l-*; both roots are clearly suppletive in synchronic terms. Some Balkan dialects (e.g. Arli of Gilan, Prilep, and Florina, Zargari, Epiros, and Rumelian Romani) possess two sets of oblique forms, a full set in *ol-* and a reduced set in *l-*.[7] In these dialects, the nominative forms and the full oblique forms share the root *o-*. Moreover, in all of these dialects but Gilan Arli, the oblique stem and the nominative plural

Table 7.17. Third-person pronouns in selected Balkan dialects

	NOM.SG: M ~ F	NOM.PL	OBL (FULL)
Crimean R	*o-v ~ o-j*	*o-n*	*l-*
Sofia Erli	*o-v ~ o-j*	*o-l*	*l-*
Arli (Gilan)	*o-v ~ o-j*	*o-n*	*o-l-*
Zargari	*o-v ~ o-j*	*o-l-a*	*o-l-*
Arli (Prilep, Florina)	*o-v ~ o-j*	*o-l-e*	*o-l-*
Epiros, Rumelian R	*o-v ~ o-j*	*o-l*	*o-l-*

form also share the segment *-l-*, and so the similarity between the case forms is even greater in the plural (e.g. Florina *ole* 'they' and *olen* 'them'). Thus, in these Balkan dialects there is no root suppletion (viz. there is a uniform root), and case irregularity is located in nominative singular inflections and in the irregular stem formative *-l-*. The Sofia Erli paradigm is more intriguing: the nominative forms share the root *o-*, and all forms but the nominative singular forms share the segment *-l-*. Nevertheless, there is clearly a synchronic root suppletion at least in the singular (e.g. *o-v* 'he' vs. *l-es-* 'him'). The person hierarchy for case differentiation is thus 3 > 1, 2.

As for **gender**, there is a clear asymmetry between the third person on the one hand, and the first and second persons on the other, both with verbs and personal pronouns. Verb forms and personal pronouns of the first and second persons generally do not encode gender; it is only in (some) third-person forms that gender is encoded. In Early Romani and in all dialects, gender is differentiated in the singular third-person pronouns (e.g. *ov* 'he' vs. *oj* 'she'), at least in the oblique cases (see Chapter 8). The only category where, in Early Romani and in some dialects, gender is encoded in verbs is the third-person singular perfective, viz. in the active participle preterite forms (e.g. *gel-o* 'he went' vs *gel-i* 'she went') and in the pluperfect forms based on the active participles (e.g. *gel-o-sas* 'he had/would have gone' vs *gel-i-sas* 'she had/would have gone'). Only some verbs may encode gender in this way (see Chapter 15).

Class differentiation in verbs has been discussed in Chapter 6. Table 6.18 showed that the differentiation of non-perfective person–number suffixes follows the hierarchy third-person singular > first-person singular > third-person plural, first-person plural > second-person singular. As for person, this hierarchy is readily interpretable as 3 > 1 > 2.

7.4. Extension

The criterion of extension renders conflicting person asymmetries. In verbs, second-person forms may extend to the third person, and vice versa, third-person forms may extend to the second person. If the first person is affected, then the extension proceeds following the scale 3 > 2 > 1. In pronouns, on the other hand, first-person forms may extend to second-person forms. The first person appears to show the greatest exposition in verbs, in that it is very infrequently affected by extensions from other persons.

In **verb** inflection, the second person and the third person interact in a conflicting way: in some dialects second-person forms (affixes or aspects of the

shape of affixes) extend to the third person, while in others the extension takes the opposite direction. Both types of extension mostly took place in the plural of (some or all) perfective sets, presumably to match, or at least to approximate, the second/third-person plural homonymy in the non-perfective sets. Less frequent extensions – those in the singular, those in non-perfective sets, and those affecting the first person – are discussed at the end of this subsection. Six patterns of marking of the second-person plural and the third-person plural perfective are shown in Table 7.18. The arrows visualise the direction of extension; double arrows indicate complete extension, simple arrows indicate partial influence (innovations in shape are marked in bold).

Type A inflections do not show any person extension. This is the original pattern of Early Romani that has also been retained in some dialects (e.g. earlier Finnish Romani, most Central dialects, Lovari and Taikon Kalderaš, and a couple of Balkan dialects). Types B and D exhibit an extension of the whole inflection, which results in homonymy between the second-person plural and the third-person plural forms. In Types C and E, there has only been a partial influence of one inflection on another, and so the relevant forms are still distinct. In Type C, the original third-person plural suffix has been influenced by the second-person plural suffix, having taken over its final consonant: *-*e* (← -*an*) > -*en*. In Type E, it was the original second-person plural suffix that has been influenced by the third-person plural suffix, having taken over its vowel quality: *-*an* (← -*e*) > -*en*. Types B and C exhibit extensions of second-person forms into the third person, and types D and E extensions of third-person forms into the second person. Type F shows mutual partial influence between the inflections. The final consonant of the original second-person plural suffix extends to the third-person plural, and the vowel quality of the third-person plural suffix extends to the second-person plural: third-person plural *-*e* (← second-person plural *-*an* or -*en*) > -*en* and second-person plural *-*an* (← third-person plural *-*e* or -*en*) > -*en*. The developments in Type F happen to result in homonymy between the second-person plural and the third-person plural forms.

Table 7.18. Extensions in second-person plural and third-person plural perfective inflections: patterns

	Type A	Type B	Type C	Type D	Type E	Type F
2PL	-*an*	-*an* ⇓	-*an* ↓		-***e***n	-***e***n ↓
3PL	-*e*		-***e******n***	-*e* ⇑	-*e* ↑	-***e******n*** ↑

The complete extension of the second-person plural inflection to the third-person plural (Type B) has occurred in most Sinti dialects (those of France, Austria, Hungary, and some of Germany), and in Roman. In Sinti the development affected both perfective sets, while in Roman it is restricted to the pluperfect. The partial influence of the second-person plural inflection on the third-person plural inflection (Type C) is found in older German Sinti, Lalere (Bohemian Sinti) and, restricted to the pluperfect, also in Welsh Romani, a couple of Central dialects, and Polish Lovari. The complete extension of the third-person plural inflection to the second-person plural (Type D) has occurred independently in a number of dialects: most Northeastern dialects (those of the Baltics, Russia, and northwest Ukraine), many current Finnish varieties, Yerli, Malokonare, Markuleš Kalderaš of Banat, and Ukrainian Vlax. The partial influence of the third-person plural inflection on the second-person plural inflection (Type E) is very common, too. It is found in Polish and Abruzzian Romani, and in the majority of the dialects of the southern Balkans (both Vlax and Balkan proper). Finally, mutual partial influence between the inflections (Type F) is found in Piedmontese Sinti and, with some variation, in Xoraxane. The dialect distribution of the types of person extension suggests that, in most instances, the extending inflection partially influences the shape of the target inflection at first (Types C and E), and only after that does it completely take over the latter's function (Types B and D). Table 7.19 illustrates each type of extension with forms of the verb *ker-* 'do, make' in selected dialects; innovative person–number suffixes are marked in bold.

The extension of third-person forms to the second person in the plural is more widespread cross-dialectally than the opposite extension.[8] However, there is another piece of evidence that suggests that the extension of second-

Table 7.19. Extensions in second-person plural and third-person plural perfective inflections: forms in selected dialects

Type	Dialect	2PL.PRET	3PL.PRET	2PL.PLPF	3PL.PLPF
A	Early Romani	**kerdj-an*	**kerd-e*	**kerdj-an-asi*	**kerd-e-s-asi*
B	Hameln Sinti	*kraj-an*	*kraj-**an***	*kraj-an-s*	*kraj-**an**-s*
	Roman	*kerč-an*	*kerd-e*	*kerč-an-ahi*	*kerč-**an**-ahi*
C	Lalere Sinti	*kerd-an*	*kerd-**en***	*kerd-an-s*	*kerd-**en**-s*
	Polish Lovari	*kerd-an*	*kerd-e*	*kerd-an-as*	*kerd-**en**-as*
D	Latvian	*kerd-**e***	*kerd-e*	*kerd-**e**-s-is*	*kerd-e-s-is*
E	Bunkuleš	*kerd-**en***	*kerd-e*	*kerd-**en**-as*	*kerd-e-s-as*
F	Piedmontese	*kerd-**en***	*kerd-**en***	*kerd-**en**-as*	*kerd-**en**-as*

person forms to the third person is equally natural. In some Sinti dialects, there is an extension of second-person forms into the third person not only in the plural (as described above), but also in the singular. First, in current German, Bohemian, Austrian, and Hungarian Sinti, the perfective second-person singular suffix *-al-* extends to the third-person singular in the pluperfect set (taking over the original third-person singular suffix *-as-* that has been retained in the preterite set). And second, in German and Austrian (but not in Bohemian or Hungarian) Sinti, the non-perfective second-person singular suffix *-eh-*, too, extends to the third-person singular in the imperfect set (taking over the original third-person singular suffix *-el-* that has been retained in the other non-perfective sets). Thus, the (rarer) extension within the singular proceeds from the second person to the third person, and never vice versa.

Finally, in a few dialects, person extensions have also affected the first person. Extension of a complete person suffix has occurred in Ukrainian dialects of Romani, both in the Podolie dialect of the Northeastern group and in the Vlax-affiliated East and West Ukrainian Romani. Here the original perfective third-person plural suffix *-e* extended not only to the second-person plural (as in the neighbouring Northeastern dialects), but also beyond, into the first-person plural. West Ukrainian Romani still retains variation in the first-person plural between the original suffix *-am* and the extended *-e*. This, too, suggests that the extension proceeded gradually: first from the third person to the second person, and only then to the first person as well. The result of the extension is a complete neutralisation of person in the plural of perfective sets in Ukrainian Romani.[9] In Abruzzian Romani, the perfective first-person plural suffix has been influenced in its vowel quality by the third-person plural suffix: **-am* (← *-e*) > *-em*.[10] Here, too, we assume a gradual extension of the vowel from the third-person plural into the second-person plural (as is attested in numerous Romani dialects, including those of Italy), and only then into the first-person plural as well.

To sum up this subsection: Person extensions in verbs may proceed in both directions between the second person and the third person (i.e. 2 > 3 and 3 > 2). None of the directions can be shown to be favoured: one of them is attested in more paradigmatic environments, while the other one is more widespread cross-dialectally and has occurred in more independent instances. first-person forms in verbs never extend to other persons, and they rarely get extended upon. If there is an extension into the first person, it proceeds gradually following the scale 3 > 2 > 1.

A piece of evidence for the extension of a first-person form (affix, in our instance) to the second person comes from the inflection of **personal**

pronouns (cf. Elšík 2000a). In Early Romani, the genitive marker *-*inř*- of the first-person singular pronoun differed from the genitive marker *-*ir*- of the second-person singular pronoun. The pattern of difference has been retained, in one form or another, in all Vlax dialects (e.g. Dasikano *m-rn*- 'my' vs. *ć-ir*- 'your.SG'), and in the North Central dialects of southern Poland and northeastern Slovakia (cf. *m-indr*- vs. *t-ir*-). In almost all non-Vlax dialects, both pronouns now agree in their genitive marking (e.g. Sepečides *m-indr*- and *t-indr*-). In some of these dialects, the uniform genitive marker can only continue the Early Romani first-person singular marker *-*inř*- (cf. Sepečides -*indr*-, Prilep Arli -*ind*-, Rumelian and Crimean Romani -*inr*-, and Razgrad Drindari -*Vř*-),[11] and so it is clear that the first-person singular marker has extended to the second-person singular. In other dialects, the uniform genitive marker contains a simple /r/ sound (e.g. -*ir*-, -*Vr*-, -*r*-), and so it could, in principle, continue either the first-person singular or the second-person singular proto-marker. However, because the simple /r/ is a regular reflex of the Early Romani cluster */*nř*/ in all of these dialects (and in no other dialects with the uniform genitive marker), it seems very likely that the extension of the first-person singular genitive marker applied here, too. We may conclude that, in some dialects certainly and in some dialects very likely, irregular genitive marking of personal pronouns has extended from the first person to the second person.

7.5. Extracategorial distribution

The third person shows the widest distribution outside of its primary categorial domain. The third person is the value that is assumed by defective modal verbs; that is most likely to be petrified in modal particles, and new infinitives; and that reflexive morphology is most likely to be modelled on. The second person outranks the first person, in that the former may be petrified in the new infinitives and model the reflexive morphology, while the latter may not. The second person is assumed by defective imperative verbs, and the first person may be petrified in modal particles.

Defective verbs are relatively uncommon in Romani. If they exist in a dialect, then they are restricted to a couple of modal and imperative verbs. Defective modal verbs only possess third-person forms. For example, in Šóka Rumungro, the modal verb *kampov*- inflects in two ways: in the debitive meaning 'ought to, should' it is fully conjugable (1a), while in the necessitative meaning 'be necessary, of need' the construction is impersonal, and the verb assumes a third-person form (1b–c). There are also slight inflectional differences between the

two meanings in some third-person forms, and so one may actually recognise two distinct verbs, a fully inflected one and a defective one.

(1) Šóka Rumungro
   a. *Kampjom onďa te džan.*
      MOD.PRET.1SG thither COMP go.INF
      'I ought to have gone there.'
   b. *Kampja mange onďa te džan.*
      MOD.PRET.3SG me.DAT thither COMP go.INF
      'I needed to go there.' (lit. 'It was necessary for me to go there')
   c. *Kampon tuke neve barāttŭa.*
      MOD.3PL you.SG.DAT new friends
      'You need new friends.' (lit. 'New friends are of need for you')

Defective imperative verbs only possess second-person forms (e.g. Šóka Rumungro *ašt-i* 'take.SG it; here it is for you.SG' and *ašt-en* 'take.PL it; here it is for you.PL'). In fact, the imperative subparadigm of any verb is defective in terms of person, possessing only the second-person forms. We may conclude that the third and the second person outrank the first person on the subcriterion of defectivity. The existence of second person defectives is clearly contingent on their imperative character.

In Early Romani, **necessity** (obligative modality) was in all likelihood encoded by inflected copula forms, which marked tense and mood and cross-referenced the subject of obligation. The proposition itself was encoded as a non-factual (subjunctive) complement of the copula (2).

(2) Early Romani (reconstructed)
   a. **On si te soven.*
      they be.3 COMP sleep.3PL
      'They have to sleep.'
   b. **Na somasi te sovav.*
      not was.1SG COMP sleep.1SG
      'I did not have to sleep.'

In some of those dialects that have retained indigenous means to encode necessity (see Chapter 14), the copula has been petrified in its original first-person singular or third-person singular present form together with the non-actual complementiser *te*, and reanalysed into an uninflected particle. The original first-person singular copula form is recognisable in the particle *humte*,

hunte, *hunde*, *unte* (< **som-te*), which is attested in the Core Sinti dialects and in Bohemian Romani (probably acquired through diffusion from Sinti). The original third-person singular copula form is found in the particle *site*, *iste* (< **si-te*) as attested in Piedmontese Sinti, the South Central dialects, Gilan Arli, Sofia Erli, Kosovo Bugurdži, Austrian Lovari and Taikon Kalderaš. As the necessity particle is uninflected, subject and TAM categories are encoded on the former complement verb (3)–(5):

(3) Bohemian Romani (Puchmayer 1821: 69)
Odoleske talan na humt- avas pheňa.
that.DAT perhaps not must be.SUBJ.1PL sisters
'For this reason we perhaps need not be sisters.'

(4) Austrian Lovari (Cech and Heinschink 1998: 72)
Iste našle -tar.
must flee.PRET.3PL away
'They had to flee.'

(5) Šóka Rumungro
Na site mange phenesahi.
not must me.DAT say.2SG.REM
'You did not have to tell me.'

The development of necessity particles from first-person singular and third-person singular copula forms testifies to a wider extracategorial distribution of first- and third-person forms, as opposed to second-person forms. While particles from the third-person singular copula have developed independently in a number of dialects, particles from the first-person singular copula are restricted to a single dialect group. The person hierarchy emerging from the development of necessity particles is 3 > 1 > 2.

In some dialects, verb forms of certain persons show extracategorial distribution in that they occur as non-finite, person-indifferent, forms in the so-called **new infinitive** constructions. This is more likely to be the case with third-person forms than with second-person forms, whereas first-person forms never develop non-finite functions. The new infinitive has developed especially in complements of modal verbs (cf. Boretzky 1996b; see also Chapter 14). In Early Romani and in most dialects (e.g. Welsh Romani, some Northwestern dialects, most Northeastern dialects, most Vlax dialects, and all Balkan dialects), modal complements are always finite: the complement verb assumes

a subjunctive form that agrees with the matrix verb in person and number. The finite subjunctive construction is illustrated from Epiros Romani (6) and Lithuanian Romani (7).[12]

(6) Epiros Romani
Dara-v-a te naš-av kokoro to skotari.
fear-1SG-PRES to walk-1SG alone at.the darkness
'I am afraid to walk alone in the dark.'

(7) Lithuanian Romani
Me dar-ow jekdžino te psir-ow de t'emnoma.
I fear-1SG one.person to walk-1SG in darkness
'I am afraid to walk alone in the dark.'

In both examples the matrix verb is in the first-person singular, and so is the complement verb. In some dialects, however, the complement verb has ceased to agree with the matrix verb. This is illustrated in examples (8)–(10):

(8) Slovak Romani (Lučivná)
Dara-f korkoro andro šišitno te phir-el.
fear-1SG alone in.the dark to walk-3SG
'I am afraid to walk alone in the dark.'

(9) Klenovec Rumungro
Dara-w korkōri te dža-n rāťaha.
fear-1SG alone to go-2/3PL with.night
'I am afraid to walk alone in the dark.'

(10) Finnish Romani (Helsinki)
Me tara-v-ā te stāv-es kokares tamlibossa.
I fear-1SG-PRES to walk-2SG alone with.darkness
'I am afraid to walk alone in the dark.'

In the three examples above the matrix verb is, again, in the first-person singular, but the complement verb takes a (historically) different person–number inflection: the (etymologically) third-person singular in Slovak Romani, the second/third-person plural in Klenovec Rumungro, and the second-person singular in Finnish Romani. The complement verb thus loses some of its finite properties: although it is homonymous to a person-inflected subjunctive form,

it is not sensitive to the person (or the number) of the matrix verb. This less finite complement form has been called the new infinitive. Historically, the new infinitive arose through petrification of a frequent person-inflected subjunctive form in the complement position.[13] The creation of the new infinitive in Romani has been triggered by contact with languages that possess an infinitive verb form.

Now, as we have seen above, the dialects that have developed the new infinitive differ with regard to which subjunctive form they petrify. The most common choice is the third-person singular: it is found in most North Central dialects (and in a couple of adjacent Rumungro varieties), in the Vendic dialects, in Slovene Romani, in Polish Romani (but no other Northeastern dialects), in the Sinti of Germany, Austria and Hungary, and in Hungarian Lovari. As a second option, most Rumungro dialects and some (non-adjacent) East Slovak dialects petrify the second/third-person plural subjunctive. Although the second-person plural and the third-person plural subjunctive forms are usually homonymous, there is some evidence that it is actually the plural of the third person rather than of the second person that the new infinitive is equated with. Šóka Rumungro, and possibly other dialects, shows the following variation in the relevant forms of middle verbs (see Section 7.2 for details): while the second-person plural form may, but need not, be contracted (e.g. *haj-on* or *haj-oven* 'you.PL understand'), the third-person plural form and, crucially, also the new infinitive are contracted obligatorily (e.g. *haj-on* but not **haj-oven* 'they understand; to understand'). A third type of the new infinitive arose from the second-person singular subjunctive form. This type is only found in Finnish Romani, and even there it alternates with a (less frequent) third-person singular infinitive.[14] Example (11a) shows the second-person singular infinitive, and example (11b) the third-person singular infinitive; both examples occurred in the speech of a single speaker.

(11) Finnish Romani (Helsinki)

a. *Phurane jūja kamm-en-a te pexx-es čēresko nāl.*
old women want-2/3PL-PRES to sit-2SG house.GEN in.front.of
'old women like to sit in front of the house'

b. *Me kamm-ā te l-el jek cāro kāli sartti.*
I want-1SG.PRES to take-3SG one mug black.F in.the.morning
'I like to have a cup of coffee in the morning.'

Finally, in Ukrainian Romani the new infinitive form is, synchronically, distinct from subjunctives of any person and number. It is marked by the suffix *-e*

with consonantal verbs (e.g. *t'er-e* 'to do' vs. *t'er-es* or *t'er-ex* 'you.SG do', *t'er-el* 's/he does' etc.), and it is markerless with vocalic verbs (e.g. *dža* 'to go' vs. *dža-s* or *dža-x* 'you.SG go', *dža-l* 's/he goes' etc.). Historically, the infinitive probably arose through erosion of second-person singular subjunctive forms as petrified in complement constructions. This is especially likely for East Ukrainian, where the second-person singular subjunctive suffix *-ex* (< **-es*) would have been already eroded to a great extent. Examples (12)–(13) show the use of the new infinitive in two varieties of Ukrainian Romani.

(12) East Ukrainian Romani (Slavjansk) (Barannikov 1934: 127)
Me kam-aw te akuš-e-pe.
I want-1SG to swear-INF-REFL
'I want to swear' (127.)

(13) West Ukrainian Romani (Kiev) (Barannikov 1934)
Na kam-l'-a lesa te džuv-e.
not want-PFV-3SG with.him to live-INF
'S/he did not want to live with him.'

To conclude, the third and the second persons clearly outrank the first person with regard to the possibility of their extracategorial distribution in the new infinitive complements: whereas the former do occur, the latter is unattested. While the use of unambiguously second-person forms is restricted to a few dialects (Finnish Romani and, probably, earlier Ukrainian Romani), unambiguously third-person forms are widespread. Moreover, there is some evidence that the person-homonymous second/third-person plural infinitives should be interpreted as third-person forms rather than second-person forms. Thus, in terms of extracategorial distribution, the third person ranks over the second person. The complete hierarchy is: 3 > 2 > 1.

The inflection of **reflexive pronouns**, which are characterised by the reflexive root *p-*, shows important parallels with the inflection of personal pronouns.[15] In various dialects we find forms that inflect parallel to third-person pronouns, and/or forms that inflect parallel to second-person pronouns and, at least at first sight, also first-person pronouns (see below). Our data reveal a clear asymmetry in person in terms of the criterion of extracategorial distribution. Table 7.20 shows oblique and genitive stems (as well as the accusative singular form) of reflexives, and the parallel forms of the relevant personal pronouns. The forms are given in their earliest reconstructable shape; most of these are attested in some dialects.

The plus-signs in Table 7.20 set off the inflections that are identical to the inflections of the relevant personal pronouns. The reflexive forms with the third-person parallelism inflect exactly like the third-person singular pronoun of the masculine gender (cf. *l-es* 'him' vs *l-a* 'her') in the singular, and like the gender-indifferent third-person plural pronoun in the plural. (The relevant inflections of the third-person pronouns are also regular nominal inflections.) All reflexive forms that show the third-person parallelism are attested.

This is not the case with reflexive forms that show the second (or first) person parallelism. First, there is no reflexive oblique singular that would parallel the second-person singular or the first-person singular pronoun (cf. *t-u-* 'you.SG' and *m-an-* 'me', but no **p-u-* or **p-an-*). And second, although the reflexive accusative singular form *pet* is clearly based on the second-person singular pronoun (it contains the accusative suffix *-t*, which otherwise only occurs in the second-person singular pronoun), the parallelism is not complete: the extant form contains a different vowel than the expected **p-u-t* (cf. *t-u-t* 'you.SG'). Historically, the reflexive form *pet* arose through an addition of the second-person singular accusative suffix *-t* to the form **pe*, which itself is a regular reflex of **pes*, i.e. of a form that shows the third-person parallelism (see above). The form *pet* thus combines or rather accumulates both parallelisms: the earlier third-person parallelism surfaces in the vowel quality, while the later second-person parallelism is reflected in the accusative suffix. The form *pet* is only attested in a couple of Rumungro varieties (as spoken, for example, in Podunajské Biskupice, Slovakia, and Salgótarján, Hungary).

The plural reflexive stems *pumen-* and *pumar-*, too, indicate that reflexive forms that are not constructed parallel to third-person pronouns follow an analogy with personal pronouns of the second rather than the first person: they contain the segment *-u-*, which is present in the second-person plural pronoun (cf. *t-u-m-en* 'you.PL') but not in the first-person plural pro-

Table 7.20. Person parallelisms in reflexive pronouns: forms

	Third person		Second [first] person	
	Reflexive	Personal	Reflexive	Personal
OBL.SG	*p+es-*	*l-es-*	–	*t-u-*
ACC.SG	*p+es*	*l-es*	*p-e+t*	*t-u-t*
GEN.SG	*p+es-ker-*	*l-es-ker-*	*p+ir-* [**p+inř-*]	*t-ir-* [**m-inř-*]
OBL.PL	*p+en-*	*l-en-*	*p+u-m-en-*	*t-u-m-en-*
GEN.PL	*p+en-ger-*	*l-en-ger-*	*p+u-m-ar-*	*t-u-m-ar-*

noun (cf. *a-m-en* 'we'). At first sight, the formation of the singular genitive stem *pir-* [**pinř-*] seems to contradict this. Dialectal forms such as *p-indr-* in Sepečides, *p-ind-* in Prilep Arli, or *p-inr-* in Rumelian Romani appear to require a reconstruction of the protoform **p-inř-*. This tentative protoform, however, would be parallel to the Early Romani first-person singular genitive **m-inř-* 'my', not the second-person singular **t-ir-*. Fortunately, there is a better explanation. All the dialects that appear to require the protoform **p-inř-* have demonstrably undergone a morphological extension of the first-person singular genitive marker **-inř-* or some reflex thereof into the second-person singular genitive (see Section 7.4). Thus, in these dialects, the formation of the reflexive genitive parallels not only the first-person singular genitive but also the second-person singular genitive (e.g. Sepečides *p-indr-* parallels both *m-indr-* 'my' and *t-indr-* 'your'). On the other hand, in those dialects that do differentiate genitive marking in the first-person singular and second-person singular pronouns, the reflexive goes together with the second-person singular and not the first-person singular pronoun (e.g. Dasikano *p-ir-* parallels *ć-ir-* 'your' but not *m-rn-* 'my'). Thus, if there was an Early Romani protoform at all (see below), it was **p-ir-* (paralleling the second-person form) rather than **p-indr-* (paralleling the first-person form). Crucially, this also means that the following generalisation is rescued: reflexive forms that are not constructed parallel to third-person pronouns follow an analogy with personal pronouns of the second person (whether or not they are also parallel to first-person pronouns).

So far we have discussed individual reflexive forms and the person parallelisms they display. Let us now look at how these forms combine into paradigms in individual dialects. Table 7.21 shows three distinct paradigm patterns with respect to person parallelism. The subtypes A2 and B2 are identical to the subtypes A1 and B1, respectively, except that they have no plural reflex-

Table 7.21. Person parallelisms in reflexive pronouns: patterns

	OBL.SG	ACC.SG	GEN.SG	OBL.PL	GEN.PL
Type A1	*pes-*	*pes*	*pesker-*	*pen-*	*penger-*
Type A2	*pes-*	*pes*	*pesker-*	–	–
Type B1	*pes-*	*pes*	*pir-*	*pumen-*	*pumar-*
Type B2	*pes-*	*pes*	*pir-*	–	–
Type C	*pes-*	*pet*	*pir-*	*pumen-*	*pumar-*

ive forms, the singular forms being used with plural antecedents as well. Slots showing the third-person parallelism are shaded.

In Type A, all available reflexive forms are constructed parallel to the third-person pronouns. This pattern is found in Welsh Romani, the Northwestern dialects, Latvian Romani, some varieties of East Slovak Romani, and some Vlax dialects (e.g. Lovari, Taikon Kalderaš, Xoraxane) (subtype A1); and in most Northeastern and most North Central dialects (subtype A2). In Type B, most reflexive forms parallel the second-person pronouns, with the exception of non-genitive singular forms (including the accusative), which show the third-person parallelism. This pattern occurs in the South Central dialects, and in most dialects of the Balkans. The South Central dialects, Gilan Arli, Sepečides, Crimean Romani, Muzikanta, Nange, and Gadžikano show the subtype B1, while Slovene Romani, Arli of Prilep and Florina, Erli, Varna Bugurdži, Malokonare, and most South Vlax dialects (e.g. Ajia Varvara, Dasikano, Varna Kalajdži, and Rešitare) show the subtype B2. As mentioned above, Type C only occurs in a couple of Rumungro varieties. This pattern has the widest distribution of the second-person parallelism, which in the accusative singular is imposed on an older third-person parallelism (see above).

To conclude our discussion on reflexive person parallelism: third-person markers are more likely to show extracategorial distribution than second-person markers in two, partly independent, respects. First, the formation of reflexive forms with the third-person parallelism is not categorially restricted, while formation of forms with the second-person parallelism is; this implies that, in the reflexive paradigm of any dialect, there is always at least one form based on the third-person pronoun (viz. the oblique singular *pes-*). And second, there are dialects where all reflexive forms show the third-person parallelism (which is *not* a logical consequence of the aforementioned), while this is never the case with the second-person parallelism (which *is* a logical consequence of the aforementioned). First-person markers may be considered to show extracategorial distribution only in case of the singular genitive *pir-* (etc.) in some dialects, and even here we have demonstrated the theoretical primacy of the second-person solution. Thus, the person hierarchy of extracategorial distribution emerging from our data is: 3 > 2 > 1.

7.6. Exposition

The first person is more exposed than the second and third persons in that it is the least likely to be affected by person extensions (see Section 7.4). Also, in

non-perfective person–number suffixes of consonantal verb classes there is a contrast between the first-person markers, which contain the vowel /a/ (< OIA */ā/: first-person singular *-av* < **-āmi* and first-person plural *-as* < **āmas*), and the second and third-person markers, which contain the vowel /e/ (< OIA */a/: second-person singular *-es* < *-asi*, third-person singular *-el* < **-ati*, and second/third-person plural *-en* < OIA **-anti*). The pattern has been inherited into Romani from Old Indo-Aryan.[16]

7.7. Borrowing

Person markers of the third person are most likely to be borrowed. Evidence from a single dialect suggests that person markers of the second person are more likely to be borrowed than person markers of the first person. The second person is also more prone to borrowing of number markers than the first person is.

Early Romani had, in all likelihood, borrowed the third-person singular present suffix *-i* from Greek. Judging from its distribution in dialects, the suffix was used with xenoclitic verbs in the so-called short non-perfective forms, while in the so-called long forms and in the imperfect as well as in oikoclitic verbs the inherited suffix *-(e)l* was retained. See Chapters 6 and 23 for the fate of the suffix in individual dialects.

In Slovene Romani, we find borrowing of other person–number markers from South Slavic (Slovene, the current L2, and/or Croatian, the previous L2). This is clearly the case with the perfective second-person plural suffix *-ate ~ -ete* (e.g. *kerdž-ate* 'you.PL did'); the non-perfective sets retain indigenous second-person plural inflections. In the first-person plural, all finite sets employ the suffix *-am-* (e.g. *ker-am* 'we do', *ker-am-a* 'we will do', *ker-am-ne* 'we were doing', and *kerdž-am* 'we did').[17] While *-am* is indigenous in the preterite, it is an innovation in the non-perfective sets (cf. Early Romani first-person plural **-as-*). One might argue that we are dealing with an internal extension from the preterite into the non-perfective sets. There is some evidence, however, that the extension has been at least facilitated, if not triggered, by contact with Slavic. In both the present-subjunctive and the preterite, there is also a first-person plural variant *-amo* (e.g. *ker-amo* 'we do' and *kerdž-amo* 'we did'), which coincides with the Slovene/Croatian present inflection *-(a)mo*. Rather than being borrowed as such, the Slavic inflection has exerted formal influence on the indigenous first-person singular suffix *-am*, and triggered or facilitated its extension from the preterite into the non-perfective sets.

Table 7.22. Person–number inflections in Slovene Romani

	1SG	2SG	3SG	1PL	2PL	3PL
PRES-SUBJ	***-o/-u***	*-e*	***-i***	***-am*(*o*)**	*-e*(*n*)	*-e*(*n*)
FUT	*-av-a*	*-eh-a*	*-el-a*	***-am****-a*	*-n-a*	*-n-a*
IMPF	*-av-e*	*-*(*v*)*s-e*	*-el-e*	***-am****-ne*	*-n-e*	*-n-e*
PRET	*-um*	*-as*	*-a*	*-am*(***o***)	***-ate***	*-e*

Finally, there is also the first-person singular suffix *-u* or *-o* in the present-subjunctive, which, again, probably arose through interaction of internal and contact induced developments. It has been claimed to be an eroded variant of the indigenous first-person singular suffix *-av* (Boretzky and Igla 1994: 393), but it is likely that the formal coincidence with the Slovene/Croatian first-person singular suffix *-u* has played a role in this development. Table 7.22 gives an overview of person–number inflections in Slovene Romani; borrowed markers or markers influenced by a contact language are in bold.

Some dialects borrow verbs from a contact language in their inflected forms, i.e. retaining the conjugation of that contact language. This is not necessarily code-switching, as the relevant contact language is not always the current L2. Numerous Romani dialects spoken in the Balkans (both Balkan and South Vlax) retain Turkish conjugation in Turkish verbs; Crimean Romani retains Crimean Tatar conjugation in Crimean Tatar verbs; and Russian and Lithuanian Romani sometimes retain Russian conjugation in Russian verbs. There appear to be no categorical restrictions with regard to the person of the borrowed inflected verb forms. However, borrowed conjugation may interact with indigenous conjugation, and this interaction reveals a person asymmetry.

In some dialects where verbs borrowed from Turkic (Turkish and Crimean Tatar) retain the Turkic conjugation, the Turkic-derived past inflections exert formal impact on the indigenous perfective inflections. This is the case in the dialects shown in Table 7.23, but not in other dialects that possess the Turkic conjugation (e.g. Muzikanta, Nange, Varna Kalajdži). Table 7.23 charts first-person plural and second-person plural perfective inflections, some of which contain the Turkic-derived plural suffix *-us* or *-əs*. It reflects the back allomorph of the Turkic suffix *-Iz*, where /I/ alternates according to rules of vowel harmony; the back allomorph is selected due to the presence of /a/ in the last syllable of the indigenous perfective forms.

In Kaspičan, Ajia Varvara and Vălči Dol, the plural suffix occurs in both persons. It is obligatory in Kaspičan (e.g. *kerdam-əs* 'we did' and *kerdan-əs* 'you.PL did'), and optional in Ajia Varvara and Vălči Dol. Evidence for person

Table 7.23. Perfective inflections containing the Turkic plural suffix *-Iz*

	1PL	2PL
Kaspičan	*-am-əs*	*-an-əs*
Ajia Varvara	*-am ~ -am-us*	*-en ~ -an-us*
Vălči Dol	*-am ~ -am-ə(s)*	*-an ~ -an-ə(s)*
Gadžikano	*-am ~ -am-əs*	*-an-əs*
Crimean R	*-am*	*-an-us*
Kalburdžu	*-am*	*-an-ə(s)*

asymmetry is found in the following three dialects. In Gadžikano, the suffix is obligatory in the second person, but only optional in the first person. And in Crimean Romani and Kalburdžu, it is restricted to the second person, while it never occurs in the first person (e.g. Crimean *kerd'an-us* 'you.PL did' vs. *kerd'am* 'we did').

In some Turkish dialects, the plural suffix *-Iz* occurs in second-person plural past forms, deriving them from the corresponding singular forms (e.g. *yaşad-ın-ız* 'you.PL lived' < *yaşad-ın* 'you.SG lived'), while first-person plural past forms take the cumulative person–number suffix *-Ik* (e.g. *yaşad-ık* 'we lived'). In other Turkish dialects and in Crimean Tatar, the separatist expression of person and number has been extended from the second-person plural to the first-person plural, and now the suffix *-Iz* marks number also in first-person plural forms (e.g. *yaşad-ım-ız* 'we lived' < *yaşad-ım* 'I lived'). Now, this person extension must have occurred independently also in some Romani dialects. In Kaspičan and Vălči Dol, verbs borrowed from Turkish consistently use the first-person plural suffix *-Ik* (e.g. *jašad-ək* 'we lived'), while other verbs consistently employ the separatist first-person plural inflection *-am-əs* (e.g. *kerd-am-əs* 'we did'). Table 7.24 shows the indigenous and Turkish-derived preterite inflections in Kaspičan.

Although we have no precise information on the Turkish dialects of Kaspičan and Vălči Dol, we feel safe in inferring that they make use of the first-

Table 7.24. Preterite inflections in Kaspičan

	1SG	2SG	3SG	1PL	2PL	3PL
Indigenous	*-om*	*-an*	*-a*	*-am-**əs***	*-an-**əs***	*-e*
Turkish-derived	*-Im*	*-In*	*-I*	*-Ik*	*-In-**Is***	*-I-ler*

person plural inflection that is retained in Turkish loans in the local Romani dialects, rather than the one used with the other verbs, i.e. *-Ik* and not *-Im-Iz*. This means that the extension of the suffix *-əs* from the second-person plural to the first-person plural is an internal innovation in the relevant Romani dialects, and that the second person is more prone to borrowing of number markers.

Chapter 8
Gender

The category of gender is, in all Romani dialects, coded in nouns, third-person pronouns, and adjectivals and, in some dialects, also in verbs. It has two values: the masculine and the feminine. Gender is an inherent (lexical) category of nouns. With adjectivals and verbs, gender is an agreement category that structures their inflectional paradigms. The status of the category with the third-person pronouns is intermediate: they show gender agreement in a wider sense (cf. Corbett 1991), and whether gender is considered to be an inherent or a paradigm-structuring category depends on how one constructs the paradigm(s) of third-person pronominal forms. Gender is cumulated with number and case in nouns, with case in adjectivals and third-person pronouns, and with person and TAM categories in verbs.

The masculine is the gender value that exhibits extracategorial distribution, and that is more likely to be renewed through borrowing or internal developments. While it is also more likely to extend to the feminine, the opposite extension is attested, too. The criteria of complexity and differentiation do not appear to assign prominence to either gender value.

8.1. Complexity and erosion

The criterion of complexity renders no unambiguous gender hierarchy. While the feminine tends to be more complex in nouns, the masculine tends to be more complex in third-person pronouns and demonstratives. There are no obvious gender asymmetries in complexity in adjectives or verbs. There is some evidence that masculine forms tend to undergo more erosion in personal pronouns.

Certain masculine **nouns**, in Early Romani and in some dialects, are zero marked in the nominative of both numbers (e.g. *vast* 'hand, hands'), while feminine nouns can be zero marked only in the nominative singular (see Chapter 6). Thus, the masculine gender tends to exhibit less structural complexity in more grammatical environments than the feminine.

A couple of dialects show a greater complexity of the masculine gender in the nominative of the third-person **pronouns**. In Early Romani and most dialects, forms of both genders are equally complex, showing a uniform root and

a monophonemic inflection (e.g. Prilep *o-v* 'he' vs. *o-j* 'she'). In some dialects, certain nominative forms of the third-person pronouns have been recently replaced by demonstrative forms (see also Chapter 12). The demonstrative extension has usually affected both genders alike (e.g. Zargari *ka-va* 'he' vs. *ka-ja* 'she', Kalburdžu *odo-va* vs. *odo-ja*). In Kaspičan, however, the demonstrative only extended to the masculine (*od-va* 'he'), while the feminine retains the old form (*o-j* 'she'). The masculine form is clearly more complex, both in its bimorphemic stem, and in its inflection. A similar development is underway in Nange, where a demonstrative is optionally used as a third-person pronoun in the masculine, but not in the feminine (cf. *ov* ~ *oda* 'he' vs. *oj* 'she').

On the other hand, the masculine forms of the third-person pronoun tend to undergo more erosion. In a couple of dialects, the masculine inflection has been eroded (e.g. **o-v* > *o*), and the masculine form now coincides with the stem, while the feminine form is overtly marked (e.g. Varna Bugurdži *o* 'he' vs. *o-j* 'she', or Sípos Rumungro *ō* vs. *ō-j*). Erosion in the third-person pronouns thus tends to lead to a lesser complexity in the masculine, i.e. opposite to the effects of morphological extensions. Nevertheless, note that the original masculine form **ov* contains two labial segments, and so it is clearly more predisposed to undergo erosion that the corresponding feminine form.

In some Central dialects (e.g. Central Slovak Romani and the Vendic dialects), the masculine nominative singular forms of **demonstratives** show a consonantal root *-d-*, while there is no consonantal root in the corresponding feminine forms: e.g. masculine *od-a* (*ad-a*) vs. feminine *o-ja* (*a-ja*) 'that' ('this'). One way this structural difference may have come about is through erosion of the root in the feminine form: cf. *oda* (< **od-ova*) vs. *oja* (< **od-ja* < **od-oja*). Nevertheless, it is also possible that forms of two originally distinct demonstrative series have been integrated into a single series. Consider the demonstrative forms of Early Romani and Central Slovak Romani in Table 8.1. According to this scenario, the least complex Early Romani demonstratives with a vocalic root have been retained in Central Slovak Romani in all forms but the masculine nominative singular.[1] On the other hand, the more complex Early Romani demonstratives (which had developed through reinforcement

Table 8.1. Demonstrative forms in Central Slovak Romani: the integration scenario

	NOM.SG.M	NOM.SG.F	NOM.PL	OBL
Early Romani	†*o-va* – *od-o-va*	*o-ja* – †*od-o-ja*	*o-la* – †*od-o-la*	*o-l-* – †*od-o-l-*
Central Slovak	*oda* (< **odova*)	*oja*	*ola*	*ol-*

by local adverbs in *Vd-*, cf. Matras 2002) have been lost in all forms but the masculine nominative singular. In other words, the demonstrative series in *od-* as such has vanished in Central Slovak Romani, only having provided a more complex form for the masculine nominative singular that has replaced the original less complex form. A similar development may also be assumed for the Vendic dialects and East Slovak Romani, where however the distribution of the more complex demonstrative forms is different. To conclude, either the feminine form has undergone more erosion than the corresponding masculine form (the erosion scenario), or the masculine form has been renewed by a structurally more complex form from a different series (the integration scenario). In either case, the masculine form is now more complex that the corresponding feminine form.

8.2. Differentiation

Like complexity, the criterion of differentiation renders no unambiguous gender hierarchy. In nouns, the feminine shows more differentiation in number, but the masculine shows more differentiation in class. Gender asymmetries with regard to differentiation in lexical type of adjectivals may assume both directions of prominence. There is no obvious gender asymmetry in inflectional differentiation in third-person pronouns, adjectivals, or verbs; and no asymmetry in case differentiation in nouns.

As mentioned above in Section 8.1, in some dialects certain masculine **nouns** show number homonymy in the nominative. The homonymy is restricted both categorially, and lexically (to a single inflectional class), and thus lack of number differentiation is not characteristic of masculines as such. Nevertheless, feminine nouns never exhibit number homonymy, and so the feminine gender may be considered to show a more systematic number differentiation. Masculine nouns, on the other hand, tend to be classified into more inflectional classes than feminine nouns. Early Romani has been reconstructed (Elšík 2000b, see also Chapter 5) as possessing eight masculine and four feminine noun classes. Dialect-specific developments retain or even increase the greater class differentiation in the masculine gender.

A gender asymmetry with regard to differentiation in lexical type or **lexicality** is less easy to evaluate. In Early Romani and in most dialects, inflection of demonstratives is distinct from that of lexical adjectives. Some dialects have assimilated certain demonstrative inflections to those of the vocalic oikoclitic class of adjectives. The assimilation, in all likelihood, occurred due to an

Table 8.2. Inflectional assimilation in demonstratives

	M.SG.NOM		F.SG.NOM	
	ADJ	DEM	ADJ	DEM
Early Romani	*-*o*	*-*V-va*	*-*i*	*-*V-ja*
East Slovak	-*o*	-*á*	-*i*	-*í*
Rumungro	-*o*	-*ā*	-*i*	-*ī*
Taikon Kalderaš	-*o*	-*o* ~ -*Vva*	-*i*	-*ja* ~ -*Vj*(*a*)
Lovari	-*o*	-*o*	-*i*	-*i*

interplay of phonological erosion of the demonstrative inflections (e.g. masculine *ad-a-va* > *ad-a* 'this') and morphological extension of the adjective inflections (e.g. *ad-a* > *ad-o*). Table 8.2 shows nominative singular inflections of adjectives and demonstratives in selected dialects that have undergone some inflectional assimilation.

In most Lovari varieties, the demonstrative and the adjective inflections are identical in both genders, and so they reveal no gender asymmetry. In some varieties of East Slovak Romani and Rumungro, the feminine inflections are segmentally identical, while the masculine inflections are different. Nevertheless, the feminine demonstrative inflections are still distinct from the ones in adjectives, in that they are stressed (East Slovak Romani) or long (Rumungro). In Taikon Kalderaš (and some other Vlax varieties), the masculine inflections are optionally identical, while the feminine inflections are different. Thus in Taikon Kalderaš, the masculine exhibits less differentiation in lexical type that the feminine, while in the previous dialects it is the other way round. It is not obvious whether it is the optional homonymy among the inflections (Taikon), or the imperfect homonymy (East Slovak and Rumungro), that should be given more weight.

8.3. Extension

Gender extension is only found in the inflection of adjectivals (including specific developments in demonstratives and the article), and in the inflection of third-person pronouns. The developments in adjectivals suggest that masculine forms will extend to the feminine, rather than vice versa. This is, however,

not confirmed by the developments in personal pronouns, where both directions of gender extension are attested.

An old, Proto-Romani, extension of masculine forms into the feminine occurred in the plural of all classes of **adjectivals** (e.g. *phur-e* 'old [PL]').[2] The lack of gender differentiation in the plural has been retained in all dialects, except for Abruzzian Romani. Only some dialects (e.g. older German Sinti, the Central dialects, Slovene Romani, Prizren Arli, Lovari, some Kalderaš varieties, and Xoraxane) possess a gender distinction in the oblique singular forms of (at least some; see also Chapter 21) vocalic adjectivals: e.g. masculine *phur-e* vs. feminine *phur-a* 'old'. This appears to be the Early Romani pattern, too (cf. Elšík 2000b). The remaining dialects (i.e. almost all northern and Balkan dialects, and South and Ukrainian Vlax) have extended the masculine inflection *-e* to the feminine as well, and so there is no gender distinction in the oblique singular forms.

As mentioned in Section 6.4, an original base (i.e. the nominative singular masculine) form of adnominal **demonstratives** may also be used with non-nominative, non-singular, and/or non-masculine heads in some dialects. If there are variant base forms, the extending variant is the one that is reduced in shape. In Polish Romani, for example, reduced demonstrative forms (e.g. *da* 'this') may be used with both genders in the singular, but the gender is kept distinct in full forms (e.g. masculine *dava* vs feminine *daja* 'this'). The development has occurred independently in a number of dialects (e.g. Welsh Romani, Finnish Romani, some Northeastern dialects, Bohemian and West Slovak Romani, and Sofia Erli). Only in Finnish Romani is the extension of masculine demonstratives into the feminine obligatory (e.g. *touva* 'that'). Optional masculine-to-feminine extensions have affected the **article** but they never lead to gender neutralisation in the singular, unless the article is completely indeclinable (as in Yerli).

Some dialects have recently lost the gender distinction in the nominative of the third-person **pronouns** due to convergence with genderless contact languages (e.g. Hungarian, Finnish, Azeri). In some Finnish Romani varieties, original masculine forms may be used in the feminine and vice versa (*jou*, *joj* 's/he'). In a few other dialects, the extension was unidirectional, whereby one of the gender forms completely replaced the other one. In Vend, Romano, some Lovari varieties, and Finnish Romani of Kuopio the original masculine forms now serve as gender-indifferent forms (e.g. Vend *ov* 's/he'), while in most Rumungro and some Lovari varieties it was the original feminine form that has extended (e.g. Šóka *ōj* 's/he'). Since both directions of the extension are equally well attested, there is no gender hierarchy.

8.4. Extracategorial distribution

Masculine markers tend to have a wider distribution than categorially appropriate. Substantival pro-words, viz. person and thing interrogatives and indefinites, as well as reflexives, generally do not inflect for gender. Nevertheless, their oblique forms contain the masculine oblique singular marker *-s-* (*-es-*, *-as-*). Thus they are constructed as if they were masculine forms. (See Section 6.5 for details.)

8.5. Internal diversity and borrowing

The criteria of cross-dialectal diversity and borrowing point to the masculine as the gender value that is more prone to renewal. Borrowing of affixes that are involved in gender marking occurred in nouns and adjectives, while there has been no borrowing of gender markers in third-person pronouns or verbs. Borrowing in adjectives does not reveal any gender asymmetry, and will not be discussed here.

After a prolonged contact with Greek, Early Romani started to borrow Greek nouns together with their nominative inflections rather than adopting their stems and adapting them by means of indigenous morphology. (Morphological adaptation was used to form non-nominative forms.) By the time of loss of contact with Greek (by most dialects), the Greek-derived nominative inflections would have been abstracted. They started to apply to post-Greek loans, constituting the basis of the xenoclitic inflectional classes (see also Chapter 23). Whereas the singular inflections have served as adaptation markers for post-Greek loans and have never been replaced by borrowing, the plural inflections have been subject to considerable renewal, both internal and contact-induced. Table 8.3 shows the Greek-derived nominative inflections, as reconstructed for Early Romani, in three major xenoclitic classes.

Table 8.3. Early Romani nominative inflections of xenoclitic noun classes

Xenoclitic classes	NOM.SG	NOM.PL
o-masculines (*Mo)	*-os*	*-i*
i-masculines (*Mi)	*-i ~ -is*	*-ja*
feminines (*Fa)	*-a*	*-es*

As for the nominative plural, the masculine xenoclitic *o*-class (*Mo) has clearly undergone most innovations. Apart from the Early Romani suffix *-i* (retained in the northern and the Central dialects, and in Slovene Romani), there is also the borrowing of Slavic *-ovi* (adapted as *-ovj-a* etc.) in some Balkan dialects of Bulgaria and Kosovo, the borrowing of Rumanian *-uri* (sometimes adapted as *-urj-a* etc.) in most Vlax dialects, and several other inflections that have resulted from internal interclass extensions. The masculine xenoclitic *i*-class (*Mi) has been much more stable, retaining the Early Romani *-ja* or its variants in most dialects (with the exception of Welsh Romani, where it has been completely replaced by the suffix *-i* of the *o*-class). The feminine xenoclitic class (*Fa) shows medium susceptability to renewal (although it is not clear how much of it is attributable to borrowing). The Greek-derived suffix *-es* is retained in most Balkan dialects. Some Balkan and Vlax dialects show *-e*, which may be either a regular reflex of the Greek suffix (e.g. in Arli, Slovene Romani, or Kalburdžu, where *-*es* > -*e*), or a loan of the South Slavic suffix *-e*. The northern and the Central dialects as well as Lovari and Taikon Kalderaš have *-i*, which in some dialects may be a borrowing of the North Slavic suffix *-i*, while in others (e.g. in the South Central dialects) it is probably an extension from the masculine xenoclitic *o*-class. Finally, some dialects that stayed in prolonged contact with Greek (e.g. Florina Arli, Sepečides, Sofia Erli, Crimean Romani, or Kosovo Bugurdži) borrowed the suffix *-Vdes* (sometimes adapted to *-Vd-a* etc.), creating another xenoclitic masculine class.

While at the Early Romani stage, inflections had been borrowed in both genders, later developments show preference for borrowing of masculine inflections (especially when one takes into account the alternative, internal, explanations for the feminine inflections). Further evidence for a greater tendency of masculine noun forms towards renewal is found in the oblique. Although there is no borrowing of inflections here, internal interclass and intraclass extensions have resulted in a greater cross-dialectal diversity in the masculine xenoclitic classes than in the feminine xenoclitic classes. Thus, for example, the oblique singular marker of the masculine *o*-class may be *-os-* or *-es-* (or both), and that of the masculine *i*-class may be *-is-* or *-es-* or *-jes-* or *-os-* (or different combinations of these), while the oblique singular marker of the feminine class is invariably *-a-* in all dialects (with the exception of Welsh Romani, where it is *-ia-*).

Chapter 9
Degree

The category of degree, which shows properties of inflection as well as derivation, exists in an overwhelming majority of Romani dialects. It is coded in adjectives and adverbs. The number of degree values differs according to dialect. Table 9.1 shows three types of the category of degree with regard to overt degree distinctions. The column headings indicate degree functions that may but need not be distinctly encoded.

The three-value paradigm exhibits maximum differentiation of overtly encoded degree values: the positive, the comparative, and the superlative. The two-value paradigm consists of the positive and a second value that covers both the comparative and the superlative functions. As a 'second degree', this value is usually termed *comparative*. However, in order to distinguish this value from the comparative proper (in the three-value paradigm), we will use the function-motivated term *non-positive*. Finally, in the third type (attested in a single dialect, viz. in modern Zargari), there is no overt category of degree. The single form, which is undifferentiated for degree, corresponds to the positive degree of the previous types in form.

The three degree types themselves reveal an asymmetry of degree functions, in terms of the criterion of exponence. The positive function is distinctly encoded in two types, while the comparative and the superlative functions are distinctly encoded in a single type (the three-value degree paradigm). Further asymmetries may be formulated over degree values (i.e. overt distinctions) rather than functions. In three-value degree paradigms, the superlative tends to be more complex than the comparative, which in turn is more complex than the positive. In two-value degree paradigms, the non-positive is more complex than the positive. Exactly the same hierarchies hold according to the criteria

Table 9.1. Types of the category of degree

	Positive	Comparative	Superlative
Three-value degree	positive	comparative	superlative
Two-value degree	positive	non-positive	
No degree	undifferentiated		

of cross-dialectal diversity and borrowing. The positive tends to be the most differentiated degree value, and positive forms may extend to the non-positive but not vice versa. Comparative forms extend more frequently to the superlative than vice versa. The hierarchy in (1) is a generalisation over the various asymmetries mentioned above.

(1) Non-positive (superlative > comparative) > positive

The higher a degree value on the hierarchy, the greater its structural complexity, cross-dialectal diversity, and susceptability to borrowing, and the lesser its exponence, differentiation, and extension. The linear ordering of degree values on the hierarchy is consistent. Some criteria construct only part of the hierarchy, but they never violate it.

9.1. Complexity

The criterion of structural complexity renders a clear degree hierarchy: the superlative tends to be more complex than the comparative, and the comparative or the non-positive is always more complex than the positive. The positive, in Early Romani and in all dialects, is the least complex degree in being consistently zero coded with regard to the other degree values (provided they are distinctly encoded). It is not necessarily zero coded in absolute terms, as it may contain overt inflectional markers (e.g. *bar-o* 'big'). Nevertheless, there is no overt degree marker in the positive.

In Early Romani, there was no morphological distinction between the comparative and the superlative functions. There was only a single non-positive form marked by the suffix *-*eder* (e.g. *bar-eder* 'bigger, the biggest'). The Early Romani pattern is continued in Welsh Romani, some varieties of Finnish Romani, some Core Sinti dialects (e.g. those of Austria), and in Russian Romani. Some other dialects (e.g. Piedmontese Sinti, Arli of Prizren and Florina, Volos Sepečides, Epiros, Rumelian Romani, North Vlax, Xoraxane, Dasikano, Priština Gurbet, and Ajia Varvara) have replaced the indigenous non-positive suffix by a loan proclitic or preposed particle, retaining the non-distinction between the comparative and the superlative functions (e.g. Kalderaš *maj baro* 'bigger, the biggest'), at least as far as specific degree marking is concerned (but see below on the grammaticalisation of the definite article). Thus, in Early Romani and in the dialects mentioned above, there can be no complexity asymmetry between the comparative and the superlative, since there

is no encoding distinction. In most dialects, however, the comparative and the superlative are distinctly encoded. There are two general patterns: either the two degrees show equal complexity, or the superlative is more complex than the comparative. In no dialect is the comparative more complex than the superlative.

First, in many Balkan and South Vlax dialects (all dialects of Macedonia and Bulgaria, Sepečides of Turkey, and Crimean Romani), both the superlative and the comparative are derived from the positive in an equipollent manner, i.e. the comparative by a specific comparative marker (e.g. Kaspičan *taa baro* 'bigger') and the superlative by a specific superlative marker (e.g. Kaspičan *en baro* 'the biggest'). In Kosovo Bugurdži, both degrees can also be formed by addition of the degree markers to the indigenous non-positive form (e.g. *po bar-eder* alongside *po baro* 'bigger', and *naj bar-eder* alongside *naj baro* 'the biggest'). Since, in these dialects, both the superlative and the comparative are derived from an identical base by overt degree markers of equal complexity (proclitics or preposed particles), there is no asymmetry between the two degrees as far as specific degree marking is concerned (again, see below on the grammaticalisation of the definite article).

In the second pattern, the superlative is derived from the comparative by means of a superlative marker: the comparative is zero coded with regard to the superlative and hence the superlative is clearly more complex. Numerous dialects have borrowed or grammaticalised specific superlative markers. In older Finnish Romani and the modern variety of Helsinki, some Core Sinti dialects, most Northeastern dialects, the Central dialects, and variantly also in Slovene Romani, a superlative marker is added to the indigenous synthetic comparative in **-eder* (e.g. Polish Romani *bar-edyr* 'bigger' > *naj-bar-edyr* 'the biggest'). In Cerhari, some Lovari varieties, and variantly also in Slovene Romani, a superlative marker is added to an analytic comparative (e.g. Cerhari *maj baro* 'bigger' > *leg-maj baro* 'the biggest'). In both instances, the degree markers occurring in the comparative are non-positive markers rather than markers of the comparative degree, since they occur in the superlative as well. The superlative markers are prefixes, proclitics, preposed particles, or preposed adjectivals (the latter, for example, in Lithuanian Romani *sam-o bar-edyr* 'the biggest'). With few exceptions (cf. the Cerhari-type of degree marking above), the superlative markers tend to show an identical or greater degree of structural independence than the comparative markers.

Some dialects where the comparative and the superlative show equal complexity in terms of specific degree marking tend to employ and grammaticalise the definite article in the superlative. This is the case both in some dialects

where the comparative and the superlative are not otherwise distinctly encoded (e.g. in Welsh Romani, Piedmontese Sinti, Florina Arli, Volos Sepečides, Ajia Varvara), and in some dialects with equipollent degree marking (e.g. in Sepečides, Nange, Malokonare, Varna Kalajdži, and Rešitare). Since the article is frequently optional in the superlative (e.g. Nange *naj o baro* alongside *naj baro* 'the biggest'), and since it may also be used with the comparative to encode definiteness of the relevant noun phrase, this criterion is less significant than the criterion of specific degree marking (see above). However, in at least some dialects (e.g. Piedmontese Sinti, Florina Arli, Ajia Varvara, Varna Kalajdži, and Rešitare), the superlative almost never occurs without the definite article. In other words, the article has been grammaticalised as part of an analytic superlative construction. For example in Varna Kalajdži, to mention just one criterion, the article is compatible with other preposed determiners in the superlative (2), which is impossible in the other degrees (3).

(2) *lesko en o cəkno čavo*
he.GEN SUP the small son
'his youngest son'

(3) *lesko* (*po*) **o cəkno čavo*
he.GEN CMP the small son
'his young(er) son'

The grammaticalisation of the definite article in the superlative renders the superlative more complex than the comparative. This is in accord with the asymmetry found in the second type of degree marking.

9.2. Differentiation

The positive tends to be more differentiated than the other degree values in two respects. First, in those dialects which retain the synthetic non-positive in **-eder*, the non-positive (or the comparative and the superlative, if they are distinctly encoded) is either indeclinable or inflects like adjectives of the consonantal class. The inflection of the consonantal class shows fewer distinctions than the inflection of the vocalic class, and so non-positive forms of adjectives of the vocalic class (which form the overwhelming majority of all adjectives) are less differentiated than the positive forms. For example, in the nominative singular, the masculine *bar-o* and the feminine *bar-i* 'big' are distinguished

in the positive, while there is a single gender-indifferent form *bar-eder* 'bigger' in the non-positive. In dialects which have lost the synthetic non-positive, there is no such asymmetry, as all degree forms retain the inflection of the positive. There is no inflectional asymmetry between the comparative and the superlative either, since both degrees always inflect in a parallel way.

Second, the positive tends to be more differentiated with regard to the word-class distinction between adjectives and adverbs. While most adverbs derived from (positive) adjectives contain an overt adverbial marker (e.g. *bar-o* 'big' > *bar-es* 'in a big way, very'), the synthetic non-positive is identical for both word-classes (e.g. *bar-eder* 'bigger; in a bigger way, more'). Again, there is no asymmetry in dialects which have lost the synthetic non-positive, as all degree forms retain the morphological potential of the positive (e.g. *po bar-o* 'bigger' > *po bar-es* 'in a bigger way'). And again, due to the morphological parallelism between the comparative and the superlative, there is usually no asymmetry between them with regard to word-class differentiation. The only exception are current German Sinti dialects, which show a greater differentiation of the superlative. Here, both word classes are homonymous in the comparative (e.g. *sik-estə* or *sik-edə* 'faster; in a faster way'), while in the superlative, adverbs may employ the loan marker *am* (e.g. *sik-estə* 'the fastest' vs. (*am*) *sik-estə* 'in the fastest way').[1]

9.3. Borrowing and internal diversity

The criteria of replicative borrowing and cross-dialectal diversity render the hierarchies superlative > comparative > positive, and non-positive > positive. The fact that the positive shows the lowest position on the hierarchies is due to the absence of overt positive markers in the relevant contact languages (so that there is no form to be replicated), and due to the consistent zero coding of the positive in Romani. Structural convergence is not indicative of any degree asymmetry.

Non-positive, comparative, and superlative markers are frequently borrowed from contact languages. Table 9.2 shows eight attested distributional patterns of borrowed degree markers. The columns represent degree functions in the recipient Romani dialect, while the values in the cells indicate the distribution of the loan degree markers in the source language.

Type I: Some dialects borrow the single degree marker that the relevant source language possesses. In both the source language and the receiving Romani dialect, the degree marker covers the comparative and the superlative

Table 9.2. Borrowed degree markers according to their functions in the L2 and in Romani

	Comparative	Superlative
Type I	non-positive	non-positive
Type II	(none borrowed)	superlative
Type III	non-positive	non-positive and superlative
Type IV	comparative	superlative
Type V	non-positive	superlative
Type VI	comparative	non-positive
Type VII	comparative	comparative
Type VIII	superlative	superlative

functions (i.e. the non-positive). This is the case of the Rumanian-derived non-positive marker *maj* in most North Vlax dialects, some western South Vlax dialects (e.g. Xoraxane, Dasikano, and Priština Gurbet), and previously also in Vidin Cocomanya, Džambazi, and Crimean Romani (see below). Further, Epiros Romani has borrowed the non-positive Albanian marker *mo* (cf. standard Albanian *më*), and Piedmontese Sinti has borrowed the non-positive Piedmontese marker *pi*. In some of these dialects, the superlative is distinguished from the comparative by the presence of a definite article (e.g. Piedmontese Sinti *pi baro* 'bigger' vs. *o pi baro* 'the biggest').

Type II: Only the superlative marker is borrowed in numerous dialects outside of the Balkans, all of which retain the synthetic non-positive in **-eder* (e.g. Polish Romani *bar-edyr* 'bigger' vs. *naj-bar-edyr* 'the biggest'). For most dialects, the superlative marker originates in their current contact languages: cf. Latvian Romani *vis-* from Latvian, Lithuanian Romani *sam-o* (an adapted form of *sam-yj*) from Russian, Hungarian Rumungro and Hungarian Sinti *leg-* from Hungarian, and Polish, West Slovak, and Slovene Romani *naj-* from Polish, Slovak, and Slovene, respectively. A few dialects retain their superlative marker from an immediately previous L2: cf. Bohemian Romani *naj-* from Moravian Czech or Slovak, and Slovak Rumungro and Roman *leg-* from Hungarian. The rare variant *naj-* in Roman is a loan from Serbian/Croatian (possibly Burgenland Croatian), an older L2. All of the relevant source languages possess, apart from the superlative marker, a non-positive marker, whose distribution and position correspond to that of **-eder* (e.g. Hungarian *nagy-obb* 'bigger' and *leg-nagy-obb* 'the biggest'). The non-positive marker usually exhibits a higher degree of allomorphy, and is less clearly segmentable, than the superlative marker. It is never borrowed as a regular degree mark-

er, and occurs only within a few lexemes (e.g. Šóka Rumungro *kíšéb* < Hungarian *késé-bb* 'later'). Thus, structural constraints on borrowing are clearly at play here, obscuring the interpretation of the categorial asymmetry.

Type III: The type II borrowing imposed on a previous type I borrowing is found in Cerhari and some Lovari varieties. Here, the Hungarian-derived superlative marker *leg-* is prefixed to the Rumanian-derived non-positive particle *maj* in the superlative (*maj baro* 'bigger' vs. *leg-maj baro* 'the biggest').

Type IV: Both a comparative and a superlative marker are borrowed in the dialects of Macedonia and Bulgaria (both Balkan and South Vlax), and in Sepečides. The relevant source languages are East South Slavic (Macedonian-Bulgarian) and Turkish. In Macedonia and Bulgaria, Turkish was, and in many areas of Bulgaria still is, the second language of Muslim Roms. Recently, it has been supplemented or replaced by the Slavic languages. Thus, Turkish is either a recent or an older current L2 for Romani dialects of Macedonia and Bulgaria, while Slavic is the (newer) current L2. For Izmir Sepečides, the only current L2 is Turkish. Table 9.3 shows various patterns in borrowing of degree markers from these two sources. The Slavic markers are the comparative *po* and the superlative *naj*. Turkish markers are represented in a uniform shape for the sake of convenience, although their precise form may vary across dialects (e.g. the comparative *daha*, *taa*, or *thaa*, and the superlative *en*, *xen*, or *em*). Rare variants are given in parentheses.

In Types A and E, both markers are borrowed from a single L2. Type E (only Slavic markers) is found in Arli of Skopje, Kumanovo, and Gilan, Sofia Erli, Yerli, Varna Bugurdži, Muzikanta, Drindari of Šumen and Razgrad, Malokonare, Kosovo Bugurdži, Rešitare, and Montana Kalajdži. The other types (B, C, D, F, and G) show a certain mixture of markers from both L2's. Types A through E conform to a generalisation (G1) that any variant of the superlative

Table 9.3. Distribution of degree markers borrowed from East South Slavic and Turkish

Dialects	Comparative	Superlative
Type A: Sepečides, Kaspičan, Gadžikano	*daha*	*en*
Type B: Kalburdžu	*daha*	*en* (*naj*)
Type C: Vălči Dol	*daha*	*naj*
Type D: Nange	*po* (*daha*)	*naj*
Type E: several dialects (see above)	*po*	*naj*
Type F: Prilep	*po*	*naj* (*en*)
Type G: Varna and Vidin Kalajdži	*po*	*en* (*naj*)

marker is no older a loan than any variant of the comparative marker. Either the former is a more recent loan than the latter (Type C), or both markers are from the same L2 (Types A and E). Type B represents a diachronic transition between Types A (Turkish) and C (mixed), and Type D represents a transition between Types C (mixed) and E (Slavic). There are two exceptions to the above generalisation, however: Types F and G. Here the older, Turkish-derived, marker is retained as a variant (a rare variant in Prilep Arli, and a basic variant in Kalajdži of Varna and Vidin) only in the superlative. Nevertheless, the generalisation holds as a strong statistical tendency, being valid for 19 out of 22 dialects. Moreover, one can also formulate an exceptionless generalisation (G2): (any variant of) the comparative marker is no younger than the superlative marker or, provided there are variant forms, than at least one variant of the superlative marker.

Types V and VI: There are also dialects where a source non-positive marker is complemented by a superlative or a comparative marker from a different L2. Through this complementation the original non-positive marker is restricted to a specifically comparative or a specifically superlative function. The markers and their L2 sources are charted in Table 9.4. Rare variants are given in parentheses.

The Rumanian-derived non-positive *maj* has been retained only in the comparative function in Vidin Cocomanya and Crimean Romani, while it has been replaced by a more recent specific marker in the superlative. In Kumanovo Gurbet of Macedonia, on the other hand, it has only been retained in the superlative, perhaps due to its formal similarity to the Macedonian-derived *naj*, while it has been completely replaced in the comparative. In Karditsa Arli, the Greek non-positive marker *pjo* is only found in the superlative. Since the Macedonian-derived comparative marker *po* is an older loan, one cannot asume a functional restriction of the original non-positive marker through a more recent borrowing as in the previous dialects. Rather, the Greek non-positive marker has been borrowed *selectively* only into the superlative func-

Table 9.4. Degree marking in dialects of Types V and VI

Dialect	Comparative		Superlative	
Cocomanya	*maj*	< Rumanian	*em* (*naj*)	< Turkish (Bulgarian)
Crimean R	*maj*	< Rumanian	*sam'i*	< Russian
Karditsa	*po*	< Macedonian	*pjo*	< Greek
Mc. Gurbet	*po*	< Macedonian	*maj*, *naj*	< Rumanian, Maced.

tion. The above generalisation G1 holds for Vidin Cocomanya (where Rumanian is an old L2, Turkish a recent L2, and Bulgarian the current L2), Crimean Romani (where Rumanian is an old L2, and Russian the current L2; Crimean Tatar, a recent L2, is not a source of degree markers), and Karditsa (where Macedonian is an old L2, and Greek the current L2). Kumanovo Gurbet is an exception here: alongside a superlative marker from Macedonian, the current L2, there is also a superlative marker from Rumanian, an old L2, while the comparative marker is only borrowed from the current L2. Again, even Kumanovo Gurbet conforms to the generalisation G2.

Type VII: In some dialects of the Balkans, a single borrowed degree marker which functions as a specific comparative marker in the source language covers both the comparative and the superlative functions in Romani. Thus in Arli of Prizren and Florina, Volos Sepečides, Rumelian Romani, and in Ajia Varvara, the Southeast Slavic comparative particle *po* is now employed as a non-positive degree marker. In some of these dialects, the superlative degree is distinguished by the presence of the definite article (e.g. Florina Arli *po baro* 'bigger' vs. *o po baro* 'the biggest'). In Ajia Varvara, the Turkish comparative particle *daha* shows the same distribution as the Slavic variant (e.g. *daha/po baro* 'bigger' vs. *o daha/po baro* 'the biggest'). The extension, in these dialects, of a specifically comparative marker into the superlative as well is due to convergence with languages that possess a single, non-positive, degree marker: Albanian in the case of Prizren Arli, and Greek in the case of the other dialects. Thus, there is an interplay of replicative borrowing and structural convergence: a specifically comparative marker was borrowed from an older L2, while the current L2 provides an innovative distribution pattern for this marker.

Type VIII: An extension in the opposite direction has occured in current German Sinti of Hameln. Here, the superlative marker *-estə*, borrowed from German *-est*(*-er*), has extended to the comparative as well (*bar-estə* 'bigger; the biggest'), paralleling the distribution of the recessive indigenous non-positive marker *-edə* (now not used in comparative constructions proper). This pattern is highly unusual in that the internally motivated extension does not render correspondence with the distribution of degree marking in the current L2 (German clearly distinguishes the comparative and the superlative). Although the German comparative marker *-er* shows similar degree of boundedness as the superlative marker, it is not borrowed into Romani.

Having considered the distributional patterns of borrowed degree markers, we are now in a position to assess the degree asymmetry between the comparative and the superlative with regard to borrowing. The type I pattern is not indicative of any asymmetry, as there is no overt distinction between the

comparative and the superlative functions. In type II, only a superlative marker is borrowed, but the emerging categorial hierarchy appears to be at least partly derivative of structural constraints on borrowing. In type VII, the extending comparative marker may have replaced a previous superlative marker, and so the fact that there is only borrowing of a source comparative marker does not provide any evidence for degree asymmetry with regard to replicative borrowing. The generalisation G2, which is formulated in terms of chronological L2 stratification and holds without exception for the types IV, V, and VI (where there are two or more borrowed degree markers), clearly shows that the superlative is more prone to renewal through borrowing than the comparative. Note also that the superlative markers in dialects of type II tend to be borrowed from current rather than older L2's. The greater susceptability to borrowing of the superlative is confirmed by the types III and VIII. Significantly, there are no dialects where only a comparative marker is borrowed and no superlative marker (or a non-positive marker borrowed selectively into the superlative function) is.

Superlative markers also show greater cross-dialectal diversity than comparative or non-positive markers. Apart from being borrowed, they can also be innovated from internal resources. Some Northwestern dialects possess the superlative marker *koni* (e.g. in Helsinki Finnish Romani) or *oni* (e.g. in Bohemian and Slovak Sinti; also *one*, probably from **koni*). The etymology of this marker is still obscure; possibly it resulted from grammaticalisation of the particle *komi* 'still', itself a loan from Greek. Most North Central dialects of Slovakia possess the superlative prefix *neg-* or *jeg-*. The former possibly results from contamination of a superlative prefix borrowed from the previous L2 (Hungarian *leg-*) by the form of the superlative prefix of the current L2 (Slovak *naj-*). The prefix *jeg-* might be a grammaticalisation of the numeral *jekh* 'one' (as reflected in its standard spelling, e.g. *jekh-bareder* 'the biggest'), or rather a contamination of the Hungarian-derived superlative prefix *leg-* by the form of the numeral. The development of the North Central superlative markers thus show an interplay of borrowing and internal innovations.

9.4. Extension

Instances of extension of degree markers or degree forms are relatively rare. The positive forms optionally extend into the comparative and superlative functions in a few dialects with an analytic non-positive (e.g. the positive *zurało mycatar* 'stronger than the cat' alongside the comparative *maj zurało*

mycatar in Ukrainian Romani). The loss of the category of degree in Zargari (see introduction to this chapter) also assumes an extension of the positive form into the other degree functions. Dialects of the borrowing type VII exemplify an extension of a comparative marker into the superlative function. The opposite extension is attested in a single dialect (type VIII).

Chapter 10
Negation

The category of negation consists of two values: affirmative and negative. It is encoded in predicate and constituent negators and the constructions they negate (e.g. verbs, modals, a few connectors, focus particles, and phasal adverbs) and in YES/NO-particles (i.e. utterance-level particles meaning 'yes' and 'no'). On one occasion, we also mention privative adjectives, which refer to the lack of quality of their base adjective. Negative indefinites have been excluded from this chapter, as negative indefiniteness is considered to be a value of the category of indefiniteness (see Chapter 19).

The negative is the more complex value (with the exception of YES/NO-particles, where there is no complexity asymmetry), which is also more likely to extend. The affirmative, on the other hand, is more diverse and more likely to be borrowed. The criterion of differentiation gives conflicting results: some negative forms are more differentiated, while others are less differentiated. There does not seem to be any negation asymmetry in erosion or extracategorial distribution.

10.1. Complexity

Affirmative elements are negated by means of a clitic or affixal negator, and they are mostly zero-coded with regard to their negative counterparts (e.g. Rumungro *kerav* 'I do' vs. *na kerav* 'I do not do'). In other words, negative elements are mostly negated affirmative elements. This is also the case with privative adjectives (e.g. *lačho* 'good' > *bi-lačho*, *či-lačho* or *na-lačho* 'evil, bad, not good'). However, in some instances to be discussed below, the difference between the affirmative and the negative is more than an addition of a negator. Negative elements of this sort are irregularly related to, or even suppletive with regard to, their affirmative counterparts.

First, some dialects of Bulgaria and Macedonia show a specific negative future due to convergence with East South Slavic. The affirmative future proclitic *ka* is replaced by a negative third-person copula form (e.g. *nanaj* in Sofia Erli and Montana Kalajdži, *nane* in Kumanovo Arli, or *naj* in Skopje Arli and

Macedonian Gurbet) plus the non-factual complementiser *te* (e.g. Sofia Erli *ka xav* 'I will eat' vs. *nanaj te xav* 'I will not eat'). In Yerli, the affirmative future proclitic *ka* requires the short (subjunctive) verb form, while the negative future is expressed through the negated long (present) verb form (e.g. *ka ovav* 'I will become' vs. *na ovav-a* 'I will not become'). In both types of the negative future, the negative value is more complex.

Second, a suppletive relation is the norm between the affirmative third-person present copula form *isi* (*esi, si, ehi, hi*) or *sin* (*hin*) 's/he is, they are' and its negative counterpart *nanaj* (*ninaj*), *nane* (*nani*) or *naj* 's/he is not, they are not'. Only in modern Sinti and, optionally, in Roman is the third-person present negative a regularly negated form (e.g. Sinti *hi nit* [is NEG], Roman *na hi* [NEG is] 's/he is not, they are not'). In all instances, the negative form is more complex than the suppletive affirmative form, as the former contains a prefix negator *na-* (*ni-*). Less frequently, suppletion also occurs between third-person past forms:[1] cf. the affirmative *isas* (*esas, sas, sys, ehas, has, his*), *isinjas* (*sinas, sina, sijas, sija*), *isine* (*sine, sne, hine*) and more 's/he was, they were' vs. the negative *nanas* (*ninas*), *nana*, or *nas* 's/he was not, they were not'. Again, the negative form contains the prefix negator *na-* (*ni-*) and tends to be more complex (e.g. Varna Kalajdži *sas* vs. *ni-nas*). However, in many dialects various developments have lead to similar surface complexity of both forms (e.g. Roman *sina* vs. *nana*, North Vlax *sas* vs. *nas*). There is also suppletion between the indigenous modal particles *šaj* 'can' and *na-šti* 'cannot'. The relationship between the modals is more regular in those dialects that possess reflexes of the Early Romani ability particle *ašti* 'can' (e.g. Welsh Romani *astis* vs. *n-astis*, Zargari *aešte* vs. *n-aešte*). Indeed, at least in some dialects, this type of ability particle appears to have been created through a secondary re-analysis of the inability particle as containing a regular verb negator (e.g. Piedmontese Sinti *stik* 'can' < *na-stik* 'cannot', Xoraxane *šti* < *na-šti*).[2] The inability particle thus tends to be more complex than its ability counterpart.[3] Third, the focus particle 'either' and the phasal adverbs 'not yet' and 'no more' are encoded as negated 'too', 'still' and 'already', respectively, in some dialects, while in other dialects they are unrelated (e.g. Bulgarian Romani *veče* 'already' vs. *po but* followed by the negator, 'no more').[4] Even so, they require a negated predicate, and so they may be considered to be more complex.

There is no obvious complexity hierarchy in YES/NO-particles. Their general shortness and frequent monosyllabicity – e.g. Arli *na* 'no', *va* 'yes' – derive from their shared discourse function.

10.2. Differentiation

The evidence for asymmetrical differentiation of affirmative and negative constructions is conflicting. On the one hand, negated verbs tend to be more differentiated in TAM categories than affirmative verbs. On the other hand, modals of affirmative ability are more likely to be differentiated than modals of negative ability.

Verb negators are commonly differentiated according to TAM categories, especially mood, of the verb they negate (e.g. Kalburdžu indicative *in* vs. subjunctive *na* vs. imperative *ma*; see Chapter 13 for details). In most dialects, *all* TAM distinctions of the negators are also made in the inflection of affirmative verbs. This means that negated verbs do not add any additional distinctions to those found in affirmative verbs. However, there are also dialects (especially Vlax but also Slovene Romani and Muzikanta) where some TAM distinctions are only made in negated verbs, through different choices of the negator, but not in affirmative verbs. This asymmetry concerns dialects where indicative vs. subjunctive negators are distinguished and where, at the same time, there is a single indicative-subjunctive verb form. For example in Lovari, the short verb form is negated by *či* in the present indicative and by *na* in the subjunctive (e.g. *či kerav* [NEG do.1SG] 'I do not do' vs. (*te*) *na kerav* '[that] I not do'). On the whole, irrespective of the above asymmetry, any dialect clearly makes more TAM distinctions in the inflection of verbs than in the negators. In other words, numerous inflectional TAM distinctions – especially aspectual and temporal ones – are not reflected in the choice of the negator. However, if we compare likes with likes, viz. affirmative verbs with negated verbs (rather than inflectional distinctions with negator selection), it becomes clear that negated verbs may be more differentiated in TAM categories than affirmative verbs. The opposite asymmetry never holds, since verb negation does not trigger any neutralisation in inflectional TAM values.[5] On the other hand, ability modals are more likely to be inflected (i.e. differentiated in person, number, and TAM) than inability modals (see also Chapter 14). This generalisation may be formulated in implicational terms: if the inability modal shows verb inflection, then the ability modal will show it as well, but not vice versa (e.g. the inflected *možin-* 'can' and the uninflected *naši* in Yerli).

10.3. Extension

Two instances of negative-to-affirmative extension are discussed below. No

extension in the opposite direction is attested. (Functional extensions in indefinites that involve negative indefiniteness will be discussed in Chapter 19.)

In Manuš, the Early Romani inability modal *našti* 'cannot' (thus retained in German Sinti) has shifted to the ability function 'can'. The Core Sinti dialects also possess the modal *naj* of obscure origin, which retains its inability function in older German Sinti, while it has shifted to the ability function in Manuš and Hungarian Sinti. Inability is marked by regular negation of the new ability modal (e.g. Manuš *našti* 'can' vs. *našti gar* 'cannot', or Hungarian Sinti *nej* 'can' vs. *nej nit* 'cannot'). Austrian Sinti represents a transitional state, where inability mostly, but not always, requires a negated *naj*. In other words, *naj* without a negator is usually interpreted as 'can' but rarely also as 'cannot'. The shift of the inability (negative ability) modals to the (affirmative) ability function has resulted from an addition of a more transparent negator and a following hyperanalysis of the inherent negative value of the modals as a property of the negator alone.

A similar process has probably occurred in the focus particle *nina* (*nin*) 'also', which is attested in Sinti, Roman, and Finnish and Latvian Romani. The particle seems to be composed of the focus particle **ni* 'also'[6] and the negator *na*, which indicates that its original meaning was the negative 'either'. Thus, if the reconstruction is correct, there must have been a negative-to-affirmative extension in the particle *nina*.

10.4. Internal diversity

Affirmative forms (YES-particles, ability modals, and volition modals) are more prone to renewal and more internally diverse than negative forms. Since there are no overt affirmative counterparts to negators ('affirmators') in Romani, the discussion of the diversity of negators is irrelevant (but see Chapter 13 for an overview of forms).

While NO-particles are reflexes of the Early Romani *na*, YES-particles show some internal diversity (as well as borrowing, see Section 10.5). At least two types of indigenous YES-particles (particles whose function it is to reply to a question in the affirmative) may be distinguished: *ova* (*uva*, *ouva*, *va*), for example in Piedmontese Sinti, West Slovak Romani, Sepečides, Kosovo Bugurdži, or Lovari; and *ehi* (*eji*, *hi*, *he*), for example in East Slovak Romani, or Arli of Prilep and Florina. Both are possibly lexicalised copula forms (cf. the 1SG subjunctive *ova* and the third-person present *ehi* < **esi*), while *ova* may be also a lexicalised demonstrative.

Modals of ability are much more diverse than modals of inability. The latter are only represented by reflexes of the Early Romani particle *našti* 'cannot' in most dialects, and the etymologically obscure form *naj* in some Sinti varieties. On the other hand, there appear to have been two indigenous ability particles already in Early Romani: *šaj* and *ašti* 'can'. The latter shows, in addition, some etymologically obscure variants: *d-ašti-* in some Kalderaš varieties and *š-ašti* (possibly a contamination of *ašti* with *šaj*) in some varieties of Lithuanian Romani. The form *saste* (*haste*) 'can' in some Sinti varieties may be related to *šašti* or, alternatively, a result of grammaticalisation of the third-person past copula form *sas* (*has*) with the non-factual complementiser *te*.[7] The form *šti* (*stik*) 'can' in Xoraxane and Piedmontese Sinti is more likely to have been created through a re-analysis of the inability particle *na-šti* as containing a regular verb negator, than through initial erosion of **ašti*. Apart from reflexes and variants of the indigenous ability particles, various dialects have (partly) grammaticalised other modal verbs into ability modals. Frequent is the use of *džan-* 'know' in this function (e.g. in Polish Romani, the Central dialects, Slovene Romani, Sepečides). Hungarian Sinti has *trun* (< *troma-* 'dare, be allowed'), and Hameln Sinti has completely replaced the indigenous ability particles by the verb *hajev-* (< **axaljov-* 'understand').

Most dialects of the Balkans have undergone or are undergoing a lexical renewal in the modal of volition: *mang-*, originally meaning 'ask for, beg', has replaced or supplemented the original *kam-* 'want'. In a few dialects (e.g. Pazardžik Malokonare, Muzikanta, and Rešitare), the replacement is restricted to affirmative contexts, while negative contexts retain the original lexeme (e.g. Malokonare *mangava* 'I want' vs. *na kamam* 'I do not want'). Thus the affirmative is more prone to internal renewal than the negative.

10.5. Borrowing

Affirmative forms are more likely to be borrowed than negative forms. This holds for YES/NO-particles, modals of ability, and inflected verb forms. The single exception to this generalisation is borrowing of a negative but no positive auxiliary in a single dialect: Slovene Romani makes use of the Slovene negative forms of the copula (e.g. *nije* 's/he is not', *niso* 'they are not') in an analytic preterite construction, while the affirmative preterite forms are indigenous (e.g. *hikle* 'they saw' vs. *niso hik* 'they did not see'). Since there are no overt affirmative counterparts to negators ('affirmators') in Romani, there

can be no asymmetry in borrowing. However, for the sake of illustration, borrowed negators are briefly discussed at the end of this section.

YES-particles are commonly borrowed (e.g. Prilep Arli and Slovene Romani *da* from South Slavic, Austrian Sinti *jo* from German, or Rumungro *igen* from Hungarian),[8] while the NO-particle is mostly the indigenous *na* (*na'a*). Only loans of emphatic NO-particles are attested (e.g. Prilep Arli *hič* 'no way; not at all' from Macedonian or Turkish). Loans of emphatic YES-particles are even more numerous, however.

Modal verbs of ability are borrowed in numerous dialects (e.g. Welsh Romani, the Northeastern dialects, Slovene Romani, and many Balkan dialects), while their negative counterparts, modals of inability, are either indigenous, or regularly negated forms of the ability loan, or both (e.g. Polish Romani *mogin-* 'can' vs. *našty ~ na mogin-* 'cannot'). Negative counterparts of other borrowed modals are regularly negated (e.g. West Slovak Romani *mosi* 'must' vs. *na mosi* 'need not'). Next, most dialects that borrow inflected Turkic verbs (see Chapters 7 and 21) negate them by means of an indigenous negator (e.g. Muzikanta *na sevejorum*, Varna Kalajdži *in severim* 'I do not like'). Only Kaspičan and Gadžikano borrow also negated Turkic forms (e.g. Kaspičan *sevmijerim* 'I do not like', a form containing the Turkish suffixal negator *-mI-*).

Borrowing of negators is rare. In current Sinti dialects, the German negators *nixt* (*nit*, *nix*) 'not' or *gar* (*gor*) 'not, not at all' have completely replaced the indigenous negator in the indicative. Finnish Romani retains the indigenous *na*, but some varieties supplement it with the zero-coded form of the Finnish negative verb *ei* 'is not, does not': the borrowed negator may be combined with the indigenous one (e.g. *ei na*, *na ei na*), or it may even occur alone (e.g. in *touva ei jānela* [that.M NEG know.3SG] 'he does not know').

Chapter 11
Cardinality

The category of cardinality is encoded in numerals. The values of the category correspond to arithmetical cardinality of the numbers the numerals designate. We employ the following terminology: order numerals are the numerals constituting the orders (levels) of the decimal arithmetical system (viz. '10', '100', '1000', etc.); unit numerals are the numerals '1' through '9'; ten numerals refer to the multiples of 10 from '20' through '90'; and ten+unit numerals are the numerals '11' through '99' with the exception of the ten numerals (i.e. '11' through '19', '21' through '29' etc.).

Higher numerals are more complex, while lower numerals are more differentiated. With regard to the criteria of internal diversity and borrowing, opposite asymmetries hold for cardinals and ordinals: while higher cardinals are more internally diverse and more likely to be borrowed from a current L2, it is lower ordinals that are more internally diverse and more prone to borrowing. Retention of older borrowed cardinals is most likely in medium cardinality. Both directions of extension are attested. Importantly, the above generalisations mostly hold only if one distinguishes different orders of the numeral system. In other words, they hold within unit numerals, within ten numerals, within ten+unit numerals, and/or within order numerals, but not necessarily if absolute cardinality is taken into account. This proviso reflects the construction of the numeral system in Romani, which is mostly decimal.

An extensive discussion of Romani numerals from a typological, and partly also historical, point of view is found in Bakker (2001). His dialect sample is larger than ours, although he did not have access to many Romani dialects (especially those spoken in the Balkans) that we include here. Unlike Bakker, we do not include Para-Romani varieties in our discussion.

11.1. Complexity

There are two structural types of numerals: simple ones and complex ones. (In this section we restrict the discussion to cardinals, since ordinals and multiplicatives are mostly derivations that retain the complexity asymmetries found in cardinals). Simple numerals consist of a single morpheme, while

complex numerals are compounds or collocations. Cardinality asymmetries in complexity derive from the fact that the numeral system of Romani is mostly decimal. Simple numerals available in most dialects are the unit numerals, the decimal order numerals, and '20'. Many dialects also possess uncompound numerals of Greek origin for tens higher than '20' (see Section 11.5 for details). The remaining cardinals are complex, being constructed by the arithmetical operations of addition, subtraction, multiplication, or division. This holds for all ten+unit numerals, higher tens, and all numerals above '100' but the order numerals. The data relevant for complexity asymmetries within individual numeral levels are discussed in greater detail in the following sections, especially Sections 11.4–5. Here we only summarise the findings for unit numerals (1) and ten numerals (2):

(1) 1–5 < 6 < 7–9
(2) 10 < 20 < Greek < higher

11.2. Differentiation

The lowest numerals – usually '1' through '3', and especially '1' – are the most differentiated: they are most likely to be inflected and show irregular derivation of ordinals and multiplicatives. The numeral '2' may show inflectional irregularity.

In Early Romani and most dialects, simple indigenous cardinals show case agreement in modifier position, inflecting like consonantal adjectives (cf. the oblique *jekh-e* '1', *trin-e* '3', *deš-e* '10' etc.).[1] Some dialects (e.g. Welsh and Finnish Romani, most Sinti varieties, Kaspičan, and Gadžikano) have lost adjectival inflection of all cardinals. More interestingly, other dialects show a split between inflecting and non-inflecting cardinals. If there is such a split, the lowest numerals tend to inflect, while higher numerals are uninflected. Thus in Rakarengo, only the numerals '1' through '3' show optional inflection, and in Šóka Rumungro, Rakitovo Yerli, and Vălči Dol, the only inflected numeral is '1'. The numeral '2' is exceptional in that it may be uninflected even if (some) higher numerals do inflect, or optionally inflected even if (some) higher numerals inflect obligatorily (e.g. in Sípos Rumungro, Velingrad Yerli, Varna Bugurdži, or Taikon Kalderaš). On the other hand, if inflected, the numeral '2' tends to show inflectional irregularity: numerous Balkan and Vlax dialects possess the oblique forms *do* (< **duj-e*) or *done*.

Regular ordinals derive from cardinals by means of the Greek-derived marker *-t(-o)*, and in some dialects, also by means of various adjectival suf-

fixes (e.g. *-itk-*, *-dir-*) or genitive forms of cardinals. The ordinal '3rd' shows a slight irregularity in many dialects: cf. *tri-to* (< *trin* '3').[2] The ordinal '2nd' is frequently regular (viz. *duj-to* < *duj* '2') but can also be suppletive, through employment of the pro-word *aver* (*vavir* etc.) 'other', or through rare loans (see Section 11.5). The ordinal '1st' is suppletive in all dialects: a few dialects retain the Early Romani form *avgo* (*vago*); numerous dialects employ the adjective *angluno* (*anglumno*, *anglanano*) 'front' or its superlative 'foremost'; and most dialects make use of borrowings (see Section 11.5). The regular ordinal *jekh-to* (< *jekh* '1') occurs only as a part of compound numerals (e.g. *deš-u-jekh-to* '11th'), or in elicitation and standardisation contexts.

Regular multiplicatives derive from cardinals by means of the indigenous marker *-var* or of a variety of borrowed or calqued markers (e.g. *-molo*, *-kopo*, *puta*, *fora*; *drom* 'times' < 'way'). Many dialects show slight irregularities for the multiplicatives of '1', '2' and/or '3' due to erosion: e.g. *jekar*, *jokhar*, *efkar*, *jefar*, *jekri* 'once' (< **jekh-var*), *du-var* 'twice' (< **duj-var*), and *trivar*, *trival*, *trijal* 'three times' (< **trin-var*). More importantly, numerous dialects (e.g. Estonian Romani, Hungarian Sinti, Sofia Erli, Kosovo Bugurdži) that have replaced the indigenous suffix by a loan/calqued marker retain the indigenous multiplicative of '1', at least as a variant. Slovene Romani multiplies '1' through '3' by means of *-(v)ar* but higher numerals by means of the loan marker *puti*. Bunkuleš Kalderaš, on the other hand, retains the indigenous suffix with all numerals but '1' (cf. *eg-data* 'once' with the multiplicative suffix from Rumanian). Both developments create irregularity of the lowest multiplicatives.

11.3. Extension

Ten+unit numerals are constructed by addition of unit numerals to ten numerals. There are two structural types of addition: by means of a connector (e.g. *deš-u-duj* '12'), and without a connector (e.g. *deš-duj* '12'). Two indigenous additive connectors are available: *-u-* (*-o-*) developed from the rarely attested conjunctive coordinator *u* 'and',[3] and *-taj-* (*-thaj-*, *-ta-*, *-te-*, *-the-*, *-ti-*) developed from the widespread conjunctive coordinator *taj* (etc.) 'and'. Apart from these, a few dialects make use of borrowed connectors (e.g. Priština Gurbet *-i-* from Serbian). There are two sorts of connector extension in the ten+unit numerals.

First, in ten+unit numerals based on simple indigenous ten numerals, Early Romani connected the indigenous unit numerals '1' through '6' by means of an overt connector, but used no connector with the Greek-derived unit numerals

'7' through '9' (e.g. *deš-u-šov* '16' vs. *deš-efta* '17'). Although such a distribution might have been originally motivated by the fact that the Greek-derived unit numerals were all vowel-initial, the phonological generalisation does not hold synchronically in some dialects (e.g. Šóka Rumungro *biš-efta* '27', but *triand-u-efta* '37').[4] While some dialects (e.g. older German Sinti, Lithuanian, Latvian and Estonian Romani, most Central and Arli dialects, Montana Kalajdži, Muzikanta, or Austrian Lovari) retain the Early Romani distribution of connectors, numerous dialects have generalised either an overt connector or no connector for all unit numerals, irrespective of their origin. If we just look at the numerals '11' through '19', Polish Romani (*deš-u-šov* '16' and *deš-u-efta* '17') and Piedmontese Sinti (*deš-ta-šov* '16' and *deš-ta-efta* '17') represent dialects that have generalised an overt connector; and Kalburdžu (*deš-šov* '16' and *deš-efta* '17') represents dialects that have generalised no connector. The developments are local (Bakker 2001), and both directions of extension are attested.

Second, some dialects show variation in connectors between ten+unit numerals based on different ten numerals. (We now disregard the variation due to the origin of the unit numerals, taking the connector with '1' to represent the whole set.) For example, Nange has three distinct connectors: *-u-* in ten+units based on '10', *-ta-* in ten+units based on '20', and zero in ten+units based on '30' (cf. *deš-u-jek* '11', *biš-ta-jek* '21', and *tranda-jek* '31'). We have unfortunately little comparative data on connectors with tens above '30'. Scarse evidence suggests that Greek-derived tens in */nda/* behave as the Greek-derived numeral *trianda* '30' (see Section 11.5); and that compound tens (see Section 11.4) either take no connector or make use of the regular conjuctive coordinator (e.g. Klenovec Rumungro *eftavārdeš efta* vs. Šóka Rumungro *eftavārdeš taj efta* '77').[5] Table 11.1 shows the various patterns.

Type A is found in Austrian and Hungarian Sinti, Manuš, and Vălči Dol;[6] Type B in Piedmontese Sinti and possibly in Sinti of Hameln; Type C1 in Finnish Romani, Lithuanian Romani, Slovak Romani of Krompachy, Nange, Rakarengo, Varna Kalajdži, and Rešitare; Type C2 in most Northeastern dialects, many North Central dialects (e.g. Bohemian, Vechec, Podhradie), Austrian Lovari, Taikon Kalderaš, Macedonian Gurbet, and Ajia Varvara; Type D1 in some Central dialects (e.g. Zborov, Klenovec Rumungro, Burgenland Roman), Yerli, Varna Bugurdži, and Crimean Romani; Type D2 in Slovak Romani of Lučivná and modern Sofia Erli; and Type E in Šóka and Sípos Rumungro and numerous Balkan dialects (e.g. Arli of Prizren, Kumanovo, Prilep and Florina, Sepečides, Rumelian Romani, Kosovo Bugurdži, Montana Kalajdži, Malokonare, and Muzikanta).

Table 11.1. Additive connectors in ten+unit numerals

	'10'	'20'	'30'
Type A	*0*	*0*	*0*
Type B	*-taj-*	*-taj-*	*-taj-*
Type C1	*-u-*	*-taj-*	*0*
Type C2	*-u-*	*-taj-*	*-taj-*
Type D1	*-u-*	*-u-*	*0*
Type D2	*-u-*	*-u-*	*-taj-*
Type E	*-u-*	*-u-*	*-u-*

The distribution of connectors across various types enables one to reconstruct the Early Romani connector *-u-* for ten+units based on '10'.[7] Types A and B then result from a secondary generalisation of no connector and the higher connector *-taj-*, respectively. Ten+units based on '30' (and other Greek-derived tens, as well as the compound tens) most commonly use the connector *-taj-* or no connector whatsoever. This suggests that the use of (reflexes of) the conjunctive coordinator *taj* 'and' is a rather late grammaticalisation development. Type E probably results from a secondary generalisation of the lower connector *-u-*. The reconstruction of the Early Romani connector used with '20' is the most intriguing. Bakker (2001: 100–101) suggests that Early Romani used *-u-* only in ten+units based on '10', while *-taj-* was employed in ten+units based on '20' through '90'. Consequently, he claims that the lower connector *-u-* is more likely to extend to higher numerals (Types D and E) than the higher connector *-taj-* is to the lower numerals (only Type B). This claim, however, crucially depends on his reconstruction. If we reconstruct the connector *-u-* for ten+units based on '20', then the higher connector *-taj-* is more likely to extend to the lower numerals (Types B and C) than the lower connector *-u-* is to higher numerals (only Type E). Whatever the reconstruction, both directions of connector extension – from lower to higher tens and vice versa – are attested, although we are not able to evaluate their relative frequency.

11.4. Internal diversity

Cardinality is clearly relevant for asymmetries in internal diversity of numerals: the higher the cardinality in cardinals and the lower the cardinality in ordi-

nals, the greater the diversity. However, in cardinals, this generalisation only holds if one distinguishes different orders of the numeral system (e.g. the order numeral '100' is much less diverse than the ten numeral '90', although it has a higher cardinality). The following are the hierarchies for unit numerals (3), ten+unit numerals (4), order numerals (5), and ten numerals (6):

(3) 1–5 < 6 < 7–9
(4) T1–T6 < T7–T9
(5) 10 (<) 100 < 1000
(6) 20 < 30 < 40–50 < 60–90

Disregarding numeral fusion (see Section 11.5), all dialects of our sample retain the Early Romani unit numerals for '1' through '6' (cf. *jekh* '1', *duj* '2', *trin* '3', *štar* '4', *pandž* '5', and *šov* '6'). Welsh Romani has two internal alternatives for '6': the divisive compound *paš durika* 'half dozen' and the multiplicative compound *duvari trin* 'twice 3'. Most dialects also retain the Early Romani numerals for '7' through '9' (cf. *efta* '7', *oxto* '8', *ennja* '9'), which had been borrowed from Greek (see Section 11.5). Two dialects of our sample show internal innovations here; the Graecisms have been completely replaced in Welsh Romani, but only supplemented in Russian Romani. The Russian Romani innovative numerals are subtractive constructions (e.g. *bi-jekh-eskiro deš* '9' = 'without-1-GEN 10'). This is also the case with the Welsh Romani *deš bi jekh* '9' (= '10 without 1'). Further innovations in Welsh Romani are additive (e.g. *trin ta štar* '7' = '3 and 4', *štar ta panš* '9' = '4 and 5') or multiplicative (*duvari štar* '8' = 'twice 4').

Ten+unit numerals are generally constructed by addition of unit numerals to ten numerals. A few dialects make use of subtractive compounds in constructing ten+units with the units '7' through '9' (e.g. Russian Romani *bi-trin-engiro trianda* '27' = 'without-3-GEN 30', alongside the additive *biš te efta* '20 and 7'). These ten+unit numerals thus show a greater diversity. The use of various additive connectors in different ten+units (see Section 11.3) does not provide a criterion for diversity asymmetry.

The numeral '10' is always *deš* (*dex*) and '100' is almost always *šel* (*xel*) or *jekh-šel* (*kšel*; = '1–100'), unless loans from the current L2 are used (see Section 11.5).[8] The numeral '1000' shows a greater diversity. Apart from numerous loans, there are a few internal formations. A few dialects (e.g. Latvian Romani, Hungarian Sinti, or some Rumungro varieties) use the multiplicative numeral *deš-šel* '10–100'.[9] The expression *baro* (*paro*, *bar*, *balə*, *bari*) in some Northeastern dialects, Piedmontese Sinti, and Abruzzian Romani

is probably based on the adjective *baro* 'big',[10] as is evidenced by the more explicit *baro šel* 'big hundred' in Ukrainian Vlax. Iranian Romani has *sile*, a loan of South Slavic *sila* 'strength, power, large amount', and its calque *zorali* 'strong one'.

Ten numerals, with the exception of the underived indigenous *biš* '20' and Greek loans (see Section 11.5), are mostly multiplicative compounds. Decimal multiplication is the norm in Romani. There are two basic structural types of decimal tens: those with the multiplicative connector *-var-* (e.g. *oxto-var-deš* '8-times-10'), and those without it (e.g. *oxto-deš-a* '8–10-PL' or *oxto-deš* '8–10'). The former type is found in Finnish Romani, Hungarian Sinti, most Northeastern dialects (e.g. Polish, Latvian, and Estonian Romani), East Slovak Romani, Rumungro, numerous Balkan dialects (e.g. all Arli varieties, Sepečides, Sofia Erli, Kosovo Bugurdži, Muzikanta, Šumen Drindari), the North Vlax, and some South Vlax dialects (e.g. Xoraxane, Dasikano, and Kosovo and Macedonian Gurbet). Hungarian Sinti and some Central dialects optionally elide the ten unit (e.g. *oxto-var* '80' = '8-times'). Prizren Arli shows a high degree of fusion of the multiplicative connector and the ten unit (viz. *-reš* < **-var-deš*), and in some tens also fusion with the unit numeral (e.g. *pa-reš* '50' < **pandž-var-deš*). Decimal tens without a connector are found in Austrian Sinti, Lithuanian Romani, Crimean Romani, and Ukrainian Vlax (plural ten unit) as well as in Welsh and Abruzzian Romani (singular ten unit). Russian Romani has variants both with and without a connector. Slovene Romani makes use of a single internally constructed ten numeral *štar-deset* '40' where the unit numeral *deset* '10' itself is a loan; other tens are direct loans (see Section 11.5). Only a few dialects (e.g. Welsh Romani and some varieties of Finnish and Russian Romani) have supplemented the indigenous *biš* '20' by a decimal multiplicative compound (e.g. Russian Romani *duj-var-deš* '2-times-10').

Two rare systems of decimal multiplication deviate from the above norm. First, a few dialects of Bulgaria (e.g. Malokonare, Nange, Varna Kalajdži, and Rešitare) construct non-borrowed tens (in practice '60' through '90') with the unit numeral '100' rather than '10' (e.g. *šov-šəl* '60' = '6–100').[11] Second, a few dialects derive some tens by the suffix *-anda* (*-eninda*), which has been abstracted from Greek loans for tens (see Section 11.5). Segmentation of the suffix was possible due to similarity of the initial parts of the Greek tens to Romani unit numerals: cf. *trin* '3' vs. *tri-anda* '30', *štar* '4' vs. *sar-anda* '40', and *p-andž* '5' vs. *p-eninda* '50'. As for derivations from indigenous unit numerals, Yerli has *š-əninda* '60' (< *šov* '6') modelled on *p-əninda*; and Abruzzian Romani has *štar-andə* '40' (< *štar* '4') modelled on *tri-andə*. Yerli also possesses

derivations from unit numerals of Greek origin. Although both parts of the derivations are of Greek origin, the forms are distinct from their Greek counterparts, and thus internal innovations: Yerli *əfta-nda* '70' (vs. Greek *evdhominda*), *oxto-nda* '80' (vs. *oghdhonda*),[12] and *əni-nda* '90' (vs. *eneninda*).

Features of vigesimal multiplication are found in some Sinti dialects and western Central dialects. Burgenland Roman allows both decimal and vigesimal compounds for all tens over '30': cf. the even ten '80' *ofto-var-deš* ('8-times-10') or *štar-val-biš* ('4-times-20'), or the odd ten '90' *eňa-var-deš* ('9-times-10') or *štar-val-biš taj deš* ('4-times-20 and 10'). Piedmontese Sinti, Manuš, and Bohemian Romani construct '40', '60' and '80' as vigesimal multiples. In West Slovak Romani, '40' and '60' are vigesimal, while '80' is decimal. The odd tens '70' and '90' are vigesimal in Piedmontese Sinti and Manuš, but decimal in Bohemian and West Slovak Romani. Hameln Sinti has vigesimal '30' *biš-ta-deš* ('20-and-10'). One dialect of the sample, Varna Bugurdži, constructs the numerals '60' and '80' as double '30' and '40', respectively (e.g. *saranda furlja* '80' = '40 pairs').

Ten numerals not constructed as multiples are much rarer. Some dialects construct '50' as 'half-100' (e.g. *paš-šel*, *jepaš-šel*), i.e. by division of a higher-level order numeral. This construction is typical of Sinti and western North Central dialects, but it is also found – alongside decimal multiples – in Welsh, Russian, and Apennine Romani. Zargari shows additive constructions for tens over '50' (e.g. *peinda jukos* '70' = '50 20'). Tens over '60' are subtractive in Russian Romani (e.g. *bi-triand-engiro šel* '70' = 'without-30-GEN 100').

Table 11.2 summarises the construction types of ten numerals, revealing the following diversity asymmetry: 60–90 > 50–40 > 30 > 20. A higher ten

Table 11.2. Construction types of ten numerals

TYPE	(SUBTYPE)	20	30	40	50	60	70–90
Simple		+	+	+	+	+	–
Multiplicative	Decimal: *-deš*	+	+	+	+	+	+
	Decimal: *-šel*	–	–	–	–	+	+
	Decimal: *-nda*	–	–	+	–	+	+
	Vigesimal	–	+	+	+	+	+
	"Pair"	–	–	–	–	+	+
Divisive		–	–	–	+	–	–
Additive		–	–	–	–	+	+
Subtractive		–	–	–	–	–	+
No. of constructions		2	3	4	4	7	7

numeral will show more diversity than, or at least as much diversity as, a lower ten numeral.

The diversity asymmetry in ordinals is '1st' > '2nd' > other (see the discussion in Section 11.2). There is no obvious diversity asymmetry in multiplicative numerals. The indigenous fraction numeral *paš* 'half' (or its derivation **jekh-paš* '1-half') is mostly retained, with the exception of a few dialects that show numeral fusion (e.g. Kaspičan, Gadžikano, Slovak Romani of Balog). There are no other indigenous simple fraction numerals. They are usually borrowed or compounded (e.g. Hungarian Sinti *firtla* 'quarter' from German, or *pānčti pāš* 'fifth' < 'fifth half/part'), thus showing greater diversity than 'half'.

11.5. Borrowing

We would like to distinguish two types of contact effects in numerals: numeral fusion with current L2's, and retention of loans from older L2's. The higher the cardinality, the more likely a cardinal from the current L2 is used. Structurally simple numerals, including order numerals, may be exempt from this rule. Cardinal loans from older L2's, on the other hand, are most likely to be retained in medium and/or order cardinalities. Unless they are fused, numerals higher than these loans are internal innovations. With ordinals, lower numerals are more likely to be borrowed than higher numerals. The evidence for cardinality asymmetries in multiplicatives is ambiguous.

The extent of numeral fusion may be great. In some dialects, only (some) simple indigenous numerals are retained. Kaspičan and Gadžikano represent the most extreme case: with the exception of '1' through '3', all numerals are borrowed from Turkish. Slovene Romani retains the indigenous '1' through '4', and only some speakers also '20'; all other cardinals are from Slovene.[13] Numerous varieties of Slovak Romani have only retained low unit numerals (e.g. '1' through '4' in Balog, or '1' through '6' in Zbojné); all others are borrowed from Slovak. In other dialects, there appears to be an arithmetic limit on non-fused numerals, irrespective of whether they are simple or compound. Thus, the North Central varieties of Podhradie and Švedlár use old (indigenous or Greek) numerals up to '29', while all higher numerals are Hungarian or Slovak, respectively. The Slovak Romani variety of Pribylina illustrates a combination of both principles: in addition to the (simple or compound) old numerals up to '29', there are also pre-Slovak forms for the simple order numerals '100' and '1000' (the latter a loan from Hungarian, the previous contact language). Numeral fusion may be, of course, gradual. Thus, Slovak Rom-

ani of the Humenné region has only old forms for '1' through '6', both old and Slovak forms for '7' through '10', '20', '100', and '1000', and only Slovak forms for all other cardinals.[14] Four Greek borrowings may be convincingly reconstructed for Early Romani: *efta* (*šta*) '7', *oxto* (*ofto*, *xto*) '8', *ennja* (*ena*, *ija*, *nja*, *ane*) '9', and *trianda* (*trenda*, *trjanada*, *tiranda*) '30'. They have no simple indigenous equivalents, and they are found in all dialects, with the exception of Welsh Romani, where they have been replaced by internal compounds (see Section 11.4). Probably Early Romani are also the Greek tens *saranda* '40' and *peninda* (*penda*) '50', which have been replaced by compounds in most dialects outside of the Balkans. Some dialects possess additional ten loans from Greek: *ikosi* (*jukos*) '20', *eksinda* '60', *evdominda* '70', *oxdonda* '80', and *eneninda* '90'. The cross-dialectal distribution of the borrowed tens shows an interesting pattern (Table 11.3).

Type A is found in most Nortwestern, most Northeastern, the western North Central, the South Central, transitional Central-Vlax (e.g. Cerhari), some North Vlax (e.g. Lovari, Taikon and Italian Kalderaš), and the northern Gurbet dialects (e.g. Srem, Bačka, Vojvodina), as well as in Abruzzian Romani. Type B is much rarer, being attested in Lombardian Sinti, Russian Romani, some North Central dialects (e.g. Krompachy), Prizren Arli, and Ukrainian Vlax. Type C is found in most Balkan and most South Vlax dialects, some North Vlax dialects (e.g. Serbian Kalderaš and Rakarengo), and in most varieties of East Slovak Romani. Type D is represented by a few South Balkan dialects (e.g. Sepečides, and Rumelian and Crimean Romani), and a few Balkan *zis*-dialects (e.g. Kalajdži of Montana and Vidin). Type E is found in Karditsa Arli,

Table 11.3. Distribution of Greek-derived ten numerals

	'20'	'30'	'40'	'50'	'60'	'70–90'
Type A	indigenous	Greek	compound			
Type B	indigenous	Greek		compound		
Type C	indigenous	Greek			compound	
Type D	indigenous	Greek				compound
Type E	indigenous	Greek				
Type F	Greek				compound	
Type G	Greek					

Type F in Iranian Romani, and Type G in Epiros.[15] The generalisation is that, above '20', borrowing of a higher ten implies borrowing of a lower ten. While tens lower than the borrowed ones are simple indigenous numerals, tens higher than the borrowed ones are invariably compounds or internal derivations. The locus of the borrowing strategy of ten numerals is '30'.

Borrowing of the Greek ten+unit numeral *dekapende* (*dekapinde*) '15' is rare: it is only attested in two Balkan *zis*-dialects, viz. Montana Kalajdži and Muzikanta. Welsh Romani possesses the loan *durika* from Greek *dhodheka* '12'.

Numerous dialects borrow the numeral '1000', which can be retained after a shift of the contact language.[16] We only exemplify dialects that retain a loan from a recent or an older L2: *xilja* (*xiles*) from Greek in Sepečides; *xiljada* (*hiljada*, *iljada*, *sijada*) from Balkan Slavic[17] in Florina Arli; *milja* (*mija*, *mia*, *mila*) from Rumanian in the South Vlax dialects, Kalderaš, and probably due to diffusion within Romani also in Kumanovo Arli, Sofia Erli, Rumelian Romani, and Šumen Drindari; *binja* (*binji*) from Turkic in Malokonare, Vidin Kalajdži, and Crimean Romani; *ezeri* (*ezeros*, *ezro*, *izero*, *sero*, *ser*) from Hungarian[18] in most Central dialects, Lovari, and and due to diffusion within Romani also in Austrian and German Sinti; *tausend* (*tojsto*) from German (e.g. in Hungarian Sinti or French Manuš); *tyśonco* from Polish in Lithuanian and Russian Romani; *ťisicos* (*tišic*) from Slovak in some North Central varieties of Czechia; and *duzštotis* from Latvian in Estonian Romani. Additional (current) loans include *tavžend* from Slovene[19] in Slovene Romani, and *hazaar* from Persian in Romano.

With ordinals, lower numerals are more likely to be borrowed than higher numerals. Only a few dialects (e.g. Russian Romani, Sofia Erli, Yerli, Varna Kalajdži) borrow the ordinal *vtoro* '2nd' from Russian or Bulgarian. Loans of the '1st' are more frequent: cf. the Slavic-derived *prvo* (*pervo*, *peršo*) in numerous dialects, the Hungarian-derived *elšēno* (or the comparative *elšebno*) in some Central dialects, and the German-derived *eršto* (*eršti*) in Sinti and Roman. The borrowed unit ordinals may be part of compound ordinals (e.g. Russian Romani *deš-u-prvo* '11th' or Šóka Rumungro *triand-u-ēšēno* '31st'). The ordinals of some higher order numerals may be in fact borrowed adjectives (e.g. Šóka Rumungro *ezerešno* '1000th' from Hungarian *ezeres* 'of 1000', rather than the regular **ezer-to*).

As discussed in Section 11.2, the lowest numerals are most likely to resist borrowing of a multiplicative marker, although there is the exception of Bunkuleš Kalderaš, which borrows a multiplicative marker precisely in the lowest numeral.

Chapter 12
Discreteness

This chapter differs somewhat from the other chapters in chapters 6–23 of the book. It deals not with a particular grammatical category, but with several different categories. What they have in common is that they all encode, in some way, the relationship between discourse presuppositions and a particular unit of information. More specifically, they help demarcate an information unit from the pool of potential units that are known or expected. We define Discreteness as the structural techniques that allow us to disambiguate chunks of information which may belong to the same or a similar information category, and to demarcate them as distinct – discrete – information units. The relevant information units may be referential entities, such as actors in thematic roles, or they may be events or actions or even entire propositional contents (full predications with modifiers and participants). The grammatical categories that are involved in marking discreteness include definite and indefinite articles, demonstratives (the Romani system is particularly complex and shows a distinction not just between situational/proximate and contextual/remote, but also between specific and non-specific), complementisers in modal constructions (which may correlate with subject-identity vs. subject-switch), complementisers in purpose clauses (which may correlate with the semantic integration of the two events), coordinating conjunctions (which highlight the degree of semantic-pragmatic integration of two clauses), and focus particles (which indicate the status of an information unit relative to expectations).

Discreteness is thus expressed through the oppositions among indexical devices and those conveying definiteness, through the use of conjunctions and other devices such as agreement or equivalent subject deletion in clause linking, through the linear ordering of constituents relative to the verb, and through the modification of propositional units by means of focus particles such as 'only' or 'even'. Within these structures, the cover-notion of discreteness might be broken down to more tailored categories such as definiteness, focus, specificity, contrast, or foregrounding.

The grammatical categories that convey stronger discreteness, disambiguation or demarcation in the sense of these individual categories (that is, more definiteness, stronger focus, contrast, etc.) tend to be structurally more complex, less prone to final-segment erosion, and in the case of free-standing

function words, more susceptible to borrowing. Structures expressing high discreteness are more likely to extend into functions with weaker discreteness. Discreteness may be given sequence priority in the linear ordering relative to the verb, so that the pre-verbal sentence positions may be used for disambiguation and demarcation purposes (foregrounding or contrast). A somewhat more ambiguous criterion with respect to discreteness is differentiation.

12.1. Complexity

Although phonology and phonetics remain largely outside the scope of our discussion, in relation to discreteness we might nevertheless mention the interplay of suprasegmental complexity, namely utterance-level stress as well as loudness, in expressing disambiguation and singling out information units as discrete and demarcated from others.

Our first morphological case for discreteness is the category of definiteness. The values here constitute a two-way dichotomy: presence of definiteness (indicated by a definite article, or another form of determiner), and absence of definiteness (indicated either by the indefinite article, or by the absence of an article). Definiteness correlates with complexity, in several ways. First, there is the straightforward and obvious complexity of nouns that are determined by an article, a demonstrative, a possessive or genitive attribute, and are thereby specified as identifiable and so as more discrete entitites. Next, there is the correlation between the complexity of indefinite nouns and their topicality status. In many dialects, use of the indefinite article indicates that a new topic is being introduced, while indefinite nouns that are not being introduced as topics remain unaccompanied by an indefinite article. Consider the following Lovari/Kelderaš examples. In example (1a), the nouns 'house' and 'pub' are used at the discourse level as attributes of the setting, and are irrelevant as topics in their own right. By contrast, in (1b) the noun 'hotel' is introduced as a topic which will continue to play a role in the progression of the discourse and the unfolding of the story:

(1) Lovari/Kelderaš (Matras 1994a: 47, 52)
   a. *Sas ame kher, muro dad puterdas kirčima.*
      was.3SG us.ACC house my father opened.3SG pub
      'We had a house, my father opened a pub.'

b. *Samas ande ek hotelo . . . taj kothe sas o Jonny maj anglal*
were.1PL in one hotel and there was.3SG ART J. more before
'We were at a hotel . . . and Jonny had been there earlier.'

An additional complexity feature of definiteness is related to case marking. Animate nouns generally tend to show accusative (oblique) marking of the direct object. However, in some dialects direct object marking may interact with definiteness. In such cases, definite animates are more likely to take oblique marking than indefinite animates. This pattern is rather common in Vlax (e.g. Kalburdžu; ex. 2), but it is found, albeit more seldom, also in other dialects (e.g. Sofia Erli; ex. 3).

(2) Kalburdžu
a. *Dikhlem ek manuš sar pirael pe ando sokako.*
saw.1SG one man how walk.3SG REFL.ACC in.ART road
'I saw a man walking down the road.' (indefinite, topical)
b. *Odova dikhla **e** pur-**e** manuš-**e.***
that saw.3SG ART.OBL old-OBL man-ACC
'He saw the old man.' (definite)

(3) Sofia Erli
a. *Me dikhljom ki ulica jekh mruš te phirel.*
I saw.1SG at street one man COMP walk.3SG
'I saw a man walking down the street.' (indefinite, topical)
b. *Ov dikhljas **e** phur-**e** manuš-**es.***
he saw.3SG ART.OBL old-OBL man-ACC
'He saw the old man.' (definite)

It seems however that oblique marking is not entirely excluded even with indefinite animates. The nominative is favoured especially when the discreteness of the direct object as a unique, identifiable entity is questioned, as in the Lovari example (4a). Oblique marking on the other hand is favoured when specificity and topicality (and so discreteness) are high, as in the Lovari examples (4b–c):

(4) Lovari

a. *Me lav mange rom.*
I take.1SG me.DAT husband
'I am getting married.' (indefinite, non-topical)

b. *Pušel jekh-**e** gaž-**es** po drom.*
ask.3SG one-OBL man-ACC on.ART road
'He asks a man on the road.' (indefinite, topical)

c. *Dikhlem **le** rom-**es.***
saw.1SG ART.OBL man-ACC
'I saw the man.' (definite)

Demonstratives outrank personal pronouns on the scale of focus: they direct attention to a given or known entity, while personal pronouns continue an already established (anaphoric) reference. Demonstratives tend to be more complex than personal pronouns in their stem composition. With the exception of shortened forms such as *do* which are in use, alongside long forms, in some of the Northeastern and Central dialects, demonstratives are almost always polysyllabic, containing even in the relatively simplex M.SG.NOM either two syllables (*kava*, *kado*, *ada* etc.) or quite often three (*adava*, *okova*, *akadava*, *adavka* etc.) or even four (*akadava*, *okodova*). Their complexity derives from the composition of consonantal stems, inflection markers, and initial vowel prefixes. Moreover, in some form or another, demonstratives almost always show a form of reduplication, either of consonantal stems (*ka-k-o*, *a-ka-v-ka*), or, more often even, of vocalic stems (*a-k-a-ja*, *o-d-o-la*, *k-a-k-a-la*, *k-o-d-o-va*, etc.). Personal pronouns, on the other hand, are almost always monosyllabic in the nominative, and in most dialects also in the markerless (independent) oblique.

Romani demonstratives are arranged in a four-term paradigm (sometimes extended through variation between short and long forms for some paradigm positions). One opposition that is encoded by the system is the presence/absence in the speech situation (also interpretable as proximity/remoteness): *kava rom* 'this man (visible in the speech situation), *kova rom* 'that man (aforementioned in the discourse context)'. The second opposition is specificity or discreteness: *kava rom* 'this man (visible)', *akava rom* 'this man (visible, and none other than he)'. In dialects that continue the Early Romani system, specificity is represented by the consonantal stems in *-k-* (*akava*, *okova*) as opposed to the plain demonstratives in *-d-* (*adava*, *odova*). But in quite a few dialects, shifts have taken place, and new forms have emerged. The forms that encode specificity are sometimes more complex than those that convey plain refer-

ence, drawing either on vocalic prefixing (specific *akava*, *adava*, plain *kava*, *dava*), or else on same-stem reduplication (specific *kakava*, plain *kadava*).

Disambiguation and contrast of reference through demonstratives involves in most dialects an interplay of vocalic and consonantal stems. While some dialects (mainly the Northwestern group and some Central dialects) show a reduced inventory of demonstratives and tend to rely on the vocalic contrast (5), elsewhere patterns of contrast may rely on an overall contrast in complexity between plain reference, and specific reference to a less accessible and so more ambiguous entity (6)–(8):

(5) Finnish Romani
Tauva *čhēr hin dārite sar* ***touva*** *varo čhēr kōri dūrite.*
this house is closer how that other house there farther
'This house is closer than that house over there.'

(6) Varna Kalajdži
Kava *stolos si kerdo kaštestar; e* ***okova*** *dikhes kote, vov si kerdo*
this chair is made wood.ABL and that see.2SG there he is made
strastəstar.
iron.ABL
'This chair is made of wood; and that one (which) you see there, it is made of metal.'

(7) Klenovec Rumungro
Adā *kher hi buteha pašeder sar* ***okodā*** *kher.*
this house is much.SOC closer how that house
'This house is much closer that that house.'

(8) Kaspičan
Kava *kher taa paši kizom* ***okaa*** *aver kute.*
this house more close how.much that other there
'This house is closer than that other one there.'

The same applies for place deixis. In Lovari, the plain forms are *kathe* 'here' and *kothe* 'there', while the specific forms are *kadka* 'precisely here' and *kudka* 'precisely there'. The following two examples from Yerli in (9) illustrate the contrast between the plain form *otka* 'there', and the more complex specific form *okutka* 'precisely there', and how the latter is employed to emphasise demarcation:

(9) Yerli

a. *Kanato uljom* ***otka*** *dikhljom či oj na ulu ləndə.*
when.REL came.1SG there saw.1SG COMP she NEG was them.LOC
'When I arrived there I saw that she was not at their house.'

b. ***Kəfka khər*** *si po pašəs otkolkoto kejka* ***okutka.***
this house is more near than that there
'This house is closer that that one over there'

Greater complexity also accompanies the representation of discrete entities in clause-linking devices. In linking main clauses with complement or purpose clauses, Romani dialects employ conjunctions (complementisers). These are sensitive to a variety of factors, among them the continuity of the subject across the two clauses. In non-factual predications, the generic Romani connector is *te* (realised as *ti* in some dialects). By contrast, the factual connector or complementiser is *KAJ* (realised as *kaj*, or replaced by a loan of the type *hoď/kə/či/oti*; see Section 12.6). The two may be combined to form a complex non-factual connector, *kaj te/hodž te* etc. The non-factual *TE* may also be modified by a (borrowed) preposition, e.g. Prilep *za te*, Epiros *ja te*, and so on.

Complementisers that link clauses whose subjects are demarcated, discrete entities (Different Subjects = DS), are more likely to be more complex than those linking clauses with Identical Subjects (IS). Table 12.1 shows the types encountered in identical-subject modal clauses with the verb 'want', compared with different-subject manipulation clauses, also with 'want'.

Most dialects seem to belong to Type A, where modal complements are generally introduced by *te*, irrespective of subject continuity. Type B, which shows optional use of the complex complementiaser, includes Bohemian Romani, Klenovec Rumungro, and Prilep Arli. Type C, showing obligatory use of the complex complementiser (either in combination with a *KAJ*-type connector, or with a preposition) in manipulation constructions, includes the dialects of Slovak Romani of Lučivná, Roman, and Polish Romani.

Marginally we also find the optional absence of a complementiser altogether. This is the case in Kaspičan and Gadžikano, where Turkish loan verbs retain

Table 12.1. Complementiser differentiation in modal and manipulative clauses

	Type A	Type B	Type C
IS with 'want' (modal)	*te*	*te*	*te*
DS with 'want' (manipulation)		*te*, *KAJ te*	*KAJ te*, PREP *te*

Turkish inflection. If a Turkish verb is used in the complement clause, the link between the clauses is conveyed by the Turkish optative (10a). Similarly, complement clauses in Šóka Rumungro may employ Hungarian(-like) infinitives with verbs either borrowed from Hungarian or derived from indigenous bases by certain Hungarian derivational markers (e.g. *huhur-āz-in-* 'collect mushrooms' < *huhur* 'mushroom') (11a). The semantically equivalent examples in each pair (10b) and (11b) contain, respectively, the complementiser and indigenous subjunctive or infinitive forms.

(10) Kaspičan and Gadžikano
  a. *Mangava odva def olsun.*
  want.1SG he disappear become.OPT.3SG(TURKISH)
  b. *Mangava odva te žal peske.*
  want.1SG he COMP go.3SG.SUBJ REFL.DAT
  'I want him to go away.'

(11) Šóka Rumungro
  a. *Gējom huhurāz-ňi.*
  went.1SG collect.mushrooms-INF(HUNGARIAN)
  b. *Gējom te huhurāzin-en.*
  went.1SG COMP collect.mushrooms-2/3PL.SUBJ(=INF)
  'I went to collect mushrooms.'

In purpose clauses, the pattern is similar in principle, though somewhat more complex. We can again distinguish several strategies, involving on the one hand the complexity of the complementiser, and on the other hand its relation to the factuality domain. The connectors employed include 1) plain non-factual *te* (or *ti*), 2) non-factual *te* preceded by a factual connector of the type *KAJ* (*kaj* or a borrowing), or by a borrowed preposition (German *um* + *te*, Greek *ja* + *te*, Serbian *za* + *te*, and so on); both types are represented in Table 12.2 as *KAJ te*, 3) the factual connector of the type *KAJ*, sometimes followed by a borrowed conjunction, e.g. Serbian/Croatian *nek*. Table 12.2 shows the various patterns found in the dialects. While four of the types show no necessary split between IS and DS constructions, in three others DS constructions are either more complex structurally, or else, as in Type F, are more closely linked to factual complements (Type A: Finnish Romani, Sípos Rumungro, Slovene Romani, Sofia Erli, Sepečides, Crimean Romani, Rumelian Romani, Šumen Drindari, Gadžikano, Kaspičan, Vălči Dol; Type B: Sinti, Lovari, Ajia Varvara, Varna Kalajdži; Type C: Klenovec Rumungro, Arli of Florina and Prilep, Varna

Table 12.2. Complementiser differentiation in purpose clauses

	Type A	Type B	Type C	Type D	Type E	Type F	Type G
IS	*te*	*te*	*te*, *KAJ te*	*te*, *KAJ te*	*KAJ te*	*KAJ te*	KAJ, KAJ *te*
DS		*te*, *KAJ te*		*KAJ te*		KAJ, KAJ *te*	

Bugurdži, Rešitare; Type D: Polish and West Slovak Romani, Šóka Rumungro, Kosovo Bugurdži, Muzikanta; Type E: Bohemian Romani, Yerli, Kalburdžu; Type F: Lithuanian Romani, Roman; Type G: Slovak Romani of Lučivná).

The comparison between Tables 12.1 and 12.2 reveals at the same time the greater complexity in the linking devices that operate among clauses that are less integrated semantically: Purpose clauses integrate two potentially independent events, while modality and manipulation predicates do not typically portray an independent action that is taken in order to achieve the goal. Some dialects indeed even distinguish between straightforward modality, as in 'must', conveying mere emotion, intention, or external force, and modality predicates such as 'try', which convey an action that is planned and actively executed:

(12) Polish Romani
  a. *Me muśindžom pšedžav pełdy vangar.*
     I must.PRET.1SG SLASP.go.1SG over coal
     'I had to go over the coal.'
  b. *Dad prubineł* ***te*** *sykaveł varyso peskre čhavenge.*
     father try.3SG COMP teach.3SG something REFL.GEN boys.DAT
     'The father is trying to teach his sons something.'

(13) Muzikanta
  a. *Pale trjabva žav ando grados.*
     again must go.1SG in.ART town
     'I must go to town again.'
  b. *O dad məčizela pes* ***tə*** *sikəl pə čavin.*
     ART father try.3SG REFL.ACC COMP teach.3SG REFL.GEN sons.ACC
     'The father is trying to teach his sons.'

These patterns follow universals of clause integration, whereby tight semantic integration correlates with plain, weaker or leass complex morphosyntac-

tic linking devices (Givón 1990; Hengeveld 1998; Cristofaro 2003). Other universal properties of event integration are also found in Romani. Disambiguation of nouns and of events can be achieved through relative clauses or adverbial subordination, respectively, both of which are complex structures in comparison with simple noun phrases or predications.

In those dialects that allow for the possibility of absence of person agreement on the complement verb ('new infinitive'), lack of agreement will depend on the continuity of the main clause subject into the complement clause. In epistemic complements, which, like purpose clauses, portray potentially independent events, the subordinated clause is almost always introduced either by a conjunction or by a pause (in a paratactic structure). In manipulation clauses, many dialects show, in addition to the complementiser, overt pronominal reference to the subject of the complement clause:[1]

(14) Finnish Romani (Helsinki)
Me kamjom ***les*** *te jal nikki.*
I wanted.1SG him.ACC COMP go.3SG away
'I wanted him to go away.'

(15) Šumen Drindari
Me mangaa ***ov*** *tə žal peske.*
I want.1SG he COMP go.3SG REFL.DAT
'I want him to go away.'

In identical-subject modality, by contrast (*I wanted to go away*), there is never repetition of the subject in the complement clause.

Finally, we might consider the remoteness marker *-as*/*-ahi*/*-s* which derives the remote tenses imperfect (from the present) and pluperfect (from the perfective) as a demarcation strategy, separating the depicted event from the context of speech (see discussion in Matras 1994a, 2001). The disambiguating category is therefore more complex than the respective context-overlapping counterpart category (the present, implying overlap with the context of speech, and the perfective, implying overlap of the outcome of an action with the context of speech).

12.2. Erosion

In the erosion of final segments of indexical expressions, those that usually express gender/number inflection, focus-carrying members of the paradigm

appear less affected than non-focus members. This can be seen in the overall tendency to maintain final -(*o*)*a*, and quite often the consonantal segments distinguishing M -(*o*)*va*, F -(*o*)*ja*, PL -(*o*)*la*, in demonstratives, against the reduction of the final vowel and the retention of just consonantal forms in the personal pronouns M -*ov*, F -*oj* etc., and finally the loss even of the consonant segment and the retention of just vocalic forms in the definite articles M *o*, F *e*/*i*, etc.

12.3. Differentiation

Demonstratives, which direct the focus of attention to a particular referent and therefore are higher on the scale of deictic focus than either personal pronouns or the definite article, are more differentiated. The minimum number of categories is two, conveying a distinction between present/proximate and absent/remote, usually expressed by means of the vocalic stem opposition -*a*- vs. -*o*-. This reduced system is predominant is varieties of the Northwestern group (Sinti and Finnish Romani), though sources from the early twentieth century still document more elaborate systems for those dialects. Three-term systems are described for a number of dialects (Piedmontese Sinti, Roman), though arguably a missing fourth member of the paradigm is simply unattested. Most dialects show paradigms of four and above. In some cases shortened forms are interchangeable with their long counterparts. As indicated above, the second semantic dimension, conveyed usually either by an opposition between consonantal stems, or by vowel or consonant reduplication, expresses specificity; see Table 12.3).

Some dialects also maintain a marginal sub-category differentiation within the set of pronouns: In Sinti and Roman, subject pronouns in *l*- are retained as highly continuous anaphora, usually in positions enclitic to the verb, sometimes enclitic to object pronouns (cf. discussion in Holzinger 1993; Matras 1999d). In other dialects, these tend to be confined to copulas or even non-verbal predications.

In line with the same tendency for maximum focus to correlate with higher differentiation, indefinite articles remain undifferentiated for gender in the nominative, contrasting with higher focus positions occupied by demonstratives, personal pronouns, and definite articles. Both definite and indefinite articles are less differentiated for case than the higher-ranking structures – demonstratives and personal pronouns. While the latter take regular nominal case-inflection patterns, the former are aligned with adjectives, showing a two-way case distinction between nominative and non-nominative. The hierarchy

Table 12.3. Demonstratives in selected dialects

Dialect	D-A	D-O	K-A	K-O
	Situation non-specific	Discourse non-specific	Situation specific	Discourse specific
Karditsa Arli	*adava*	*odova*	*akava*	*okova*
Prilep Arli	*adaa*	*odoa*	*akaa*	*okoa*
Florina Arli	*adava*	*odova*	*akava*	*okova*
Slovene R	*dava*	*dova*	*kava*	*kova*
Nógrád Rmg	*adā*	*odā*		*okova*
Epiros	*ava*	*ova*	*akava*	*okuva*
Sinti			*kava*, *kau*	*kova*, *ko*
Finnish R	*dauva*, *dai*	*douva*, *doi*		
Lithuanian R	*dava*, *da*	*dova*, *do*		
Klenovec Rmg	*adā*	*odā*	*akadā*	*okodā*
Šóka Rmg	*adā*, *āk-adā*	*odā*, *ōk-odā*	*akā*, *āk-akā*	*okā*, *ōk-okā*
Lovari	*kado*	*kodo*	*kako*	*kuko*

for differentiation is thus: demonstratives > personal pronouns > definite articles > indefinite articles.

12.4. Extension

In the domain of indexical expressions, categories with stronger focus properties regularly extend into categories with weaker focus. The most widespread set of personal pronouns, those deriving from Early Romani *ov*/*oj*/*on*, can be derived from Proto-Romani remote demonstratives **ova*/**oja*/**ola* (Matras 2002, Ch. 5). The set of definite articles can in turn be derived from the same set of personal pronouns. The pattern for extension is thus: demonstrative → personal pronouns → definite article, matching the hierarchy of focus. Extension processes of demonstratives to personal pronouns are still active in individual dialects, and we find them in various varieties of northern Bulgaria (Gadžikano, Nange, Muzikanta, Kaspičan, Malokonare, and Kalburdžu *odva*, *odova*, or *oda*), and in Kuopio Finnish Romani *douva*, Zargari *kava*, and Abruzzian plural *kula*.

Another extension path of demonstratives is into the category of fillers and tags, where the remote demonstrative (often *kova*) may be used in the sense of 'and so on', or as a pause-filler (e.g. Lovari *kuko*, Dasikano *gua*, both search

tokens). The specific demonstrative is often lexicalised to mean 'the other one', as in *ma asa okole dženenge* 'don't laugh at other people'. It also specialises in expressions of remote time: Yerli *akava kurko* 'last week', Lovari *kuko berš* 'last year', Varna Kalajdži *po okola gijes* 'the day before yesterday' and *po akola gijes* 'the day after tomorrow', Rešitare *okoja rat* 'last night'.

12.5. Exposition

Demonstratives are exposed, in comparison with other referential categories such as personal pronouns and definite articles, in showing unique gender/number inflection markers. The Early Romani forms, which continue in most dialects, are M *-va*, F *-ja*, PL *-la* (for the origin see Matras 2002: 106–108). The M is often weakened to *-a*, and some dialects extend partly nominal inflection patterns to demonstratives.

12.6. Internal diversity and borrowing

Indexical devices that encode definiteness (articles and demonstratives) are extremely diverse across the dialects. In demonstratives, this is an outcome of the frequent need to reinforce deictic focus through combinations with other deictic stems, or with place deictics (thus Early Romani *adaj*, *akaj* 'here, precisely here' and *odoj*, *okoj* 'there, precisely there' in combination with *-va* give rise to *akava*, *adava* etc.; cf. Matras 2002, Ch 5). Among definite articles, diversity may be attributed to dialect-specific processes of erosion (see above). In the demonstrative set, loss of the four-term system in some dialects usually results in the neutralisation of the opposition specific:plain (see Table 12.3). Although elements that are high on the scale of focus appear more resistent to erosion, in this case we might attribute the loss of differentiation to contact influence and convergence with the system of demonstratives of the contact language; thus in the case of (recent) developments in the Northwestern group. Definite articles are also prone to contact, though not to borrowing of forms, but to convergence of patterns. The emergence of the preposed definite article in Early Romani is the outcome of convergence with Greek, and definite (and in some cases indefinite) articles disappear in some dialects in contact with Slavic languages, notably Slovene Romani and the Northeastern group.

Discreteness categories, in particular contrast and focus, and the linking of independent events, correlate directly with susceptibility to borrowing in free-

standing function words (conjunctions and focus particles), leading in turn to a high diversity of the respective forms among the dialects. All dialects of Romani borrow the conjunction 'but' from a current or recent contact language (e.g. Slavic *no*, *po* and *ali*/*ale*, Hungarian *de*, Turkish *ama*, German *aber*). A pre-European expression for 'or', *vaj*, is retained in some dialects (Welsh Romani, North Vlax, North Central). Elsewhere, 'or' is often found to be a more stable borrowing than 'but': Ajia Varvara for instance has *ja* 'oder' from its Recent L2 Turkish, but *ala* 'but' from its Current L2 Greek, Helsinki Romani has *elle* 'or' from its Recent L2 Swedish, but *mut* 'but' from its Current L2 Finnish. Pre-European *ta*(*j*) 'and' is retained in well over half the dialects in the sample, though often alongside a borrowed conjunction for 'and'. The additive conjunction too may be a more conservative, earlier loan than the contrastive or alternative conjunction; thus Bugurdži *hem*/*em* from the recent L2 Turkish, alsongside pre-European *thaj*, but *ili* 'or' and *po*/*ali* 'but' from the Current L2, Serbian. There is a clear implicational hierarchy for the borrowing of coordinating conjunctions, based on contrast (cf. Matras 1998b, 2002): 'but' > 'or' > 'and'.

The contact-susceptibility of contrastive conjunctions is confirmed by the overwhelming borrowing of concessive 'although' from the Current or Recent L2s, with few exceptions (Slovak Romani *the te* lit. 'also if', Lovari *sa jekh* 'despite', lit. 'all one', Prilep *i te*, combining borrowed *i* 'and' with inherited *te* 'if', Roman *kajk kaj*, the first component being an original indefinite **kaj-jekh* 'some'). Concessive condition is usually expressed by a borrowed focus marker meaning 'even', and the conditional particle *te* (or a borrowing).

Focus particles, which single out discrete information units, equally support the impression that contrast favours borrowability. Here, the restrictive particle 'only' is at the top of the scale, all dialects employing a form from the Current or Recent L2 (e.g. Hungarian-derived *čak*, Turkish *sade*, South Slavic *samo*, Polish *tylko*, German *nur*, Rumanian *numa*/*feri*). It is followed by 'even', which is also normally borrowed (e.g. Russian *daže* and *xot'*, Bulgarian *makar* and *dori*, Hungarian-derived *mek*), though some dialects show combinations with inherited items, often a conditional marker (*mek te*, *dori te*, *ili te*), or even pre-European *vi* 'also'. More conservative is 'too', for which dialects often have pre-European *vi*, *te*, or in the Northwestern group *nina*. Nonetheless, around half of the sample dialects show borrowings (sometimes conservative borrowings) for 'too' (e.g. Turkish *da*, Hungarian *iš*, Russian *tože*, West Slavic *tiš*/*tyš*, South Slavic *i*). The implicational hierarchy for borrowing is: 'only' > 'even' > 'too'.

Complex connectives, finally, show a related hierarchy. At the top of the borrowability scale for these we find the negative 'neither – nor', which is

always a borrowing (often Slavic *ni – ni* or *ani – ani*, or Turkish *ne – ne*). For 'either – or', some dialects employ inherited *vaj – vaj* (North Vlax, Central), while others often retain a borrowing from a Recent rather than Current L2 (Helsinki Romani *elle – elle* from Swedish, Kalburdžu *ja – ja* from Turkish). For 'both – and', most dialects draw on inherited forms such as *ta* 'and', *sar – ta* lit. 'how – and', *te – te*, *vi – vi*, or *o duj* 'the two'. There are also numerous borrowings (Slavic *i – i*, *ili – ili*, Turkish *da – da*), once again often retentions from an earlier L2 (e.g. Sofia Erli *hem – hem*). Once again there are no exceptions to the implicational hierarchy for borrowing: 'neither – nor' > 'either – or' > 'both – and'.

Borrowing of complementisers correlates in a comparable way with discreteness, in that complements that introduce potentially independent events or topics and so are markers of greater discreteness are more likely to be borrowed. As discussed above, complements that convey potentially independent events such as purpose clauses, epistemic complements, and different-subject modality clauses (manipulation) are more likely to take a conjunction of the type KAJ. In roughly half of the sample dialects, the Early Romani form *kaj* is retained (or replaced by inherited *so* in some of the Northeastern dialects). Elsewhere, it is found alongside a borrowing, or has been fully replaced by a borrowing. Borrowings include Rumanian *kə/ke* in Vlax, Hungarian *hoď/hod/hoj* primarily in the Central dialects, Greek *oti*, Bulgarian *či/če*, Swedish-derived *at* in Finnish Romani, Balkan Turkish *ani* (< *hani* 'where') in some of the dialects of northern Bulgaria. The non-factual counterpart *te/ti* is never replaced by a borrowing. The one exception is the Slovene Romani dialect, though here it appears that the two complemetisers merged in **ti* in a similar development as attested in Sinti and Welsh Romani, before the L2-form was adopted; this is supported by evidence from the closely related variety in Istria, where *ti* is the general complementiser, factual and non-factual (cf. Matras 2002: 210).

12.7. Linear order

In most dialects, the appearance of the subject in pre-verbal position is used to establish the subject as the specific point of departure in the information structure of the sentence. This can be seen as one of two ways of achieving coherence in discourse: through juxtaposition of subject-actors (SV), or through the chaining of predicates (VS). The juxtaposition of actors places the discreteness of subject-actors in the foreground, and is 'categorical' in Sasse's (1995)

terms, while the chaining of predicates puts the connections among predicates in the foreground, and constitutes according to Sasse a 'thetic' relation:

(16) Sofia Erli
a. *'Pokhinav' phenen o Vlahja; o Erlides phenen 'khinav'.*
pay.1SG say.3PL ART Vlachs ART Erli say.3PL pay.1SG
'*pokhinav* is what the Vlach say; the Erli say *khinav*'.
b. *Savore zanakle, na zanaklo o phuro.*
all SLASP.left.3PL NEG SLASP.left.3SG.M ART old.M
'Everybody left, [but] the grandfather did not leave.'

(17) Lovari/Kelderaš (Matras 1994a:122):
Miri mami mindig phenelas/ pušos latar:
my grandmother always say.3SG.REM ask.1SG.REM her.ABL
"Sostar phenen amenge, e romenge, 'čor'"
why say.3PL us.DAT ART Roma.DAT thief
A phenel muri mami: "Kodo sas kade: . . ."
and say.3SG my grandmother that was such
'My grandmother always used to say/ I used to ask her:
"Why do they call us, the Roma, *thieves*?"
And my grandmother says: "It was like this: . . ."'

SV word order thus operates on the basis of the discreteness of the subject-entity, relying on it in order to establish the perspective of the sentence. In opposition to the connective or consecutive VS order, SV serves as a strategy of foregrounding subjects. In some dialects, the postpositioning of attributes may be used as a disambiguation strategy, placing emphasis on the discreteness of the head by means of exposing the attribute.

Chapter 13
Tense, aspect, and mood

In this chapter we discuss inflectional TAM categories (tense, aspect, and mood) and also, rather marginally, aktionsart or actionality categories (especially iterativity). Tense, aspect and mood functions do not combine in a completely transparent way. Combinations of TAM functions that are, in some or all dialects, encoded in distinct sets of forms constitute the language-specific TAM values. The TAM paradigm of the copula is partly different from the TAM paradigm of lexical verbs (see below).

Using traditional labels, the functions of individual TAM values in lexical verbs may be described as follows. The *present* encodes indicative present and the *future* encodes indicative future. The *imperfect* encodes imperfective past indicative, or real/potential conditional. The *preterite* (also called aorist) encodes perfective past indicative. The *pluperfect* encodes the pluperfect and/or unreal conditional. The *subjunctive* encodes present subjunctive, and the *imperative* encodes present imperative. Aspect is encoded only in the past sets. Matras (2001, 2002) proposed two analytic innovations based on the assumption of form–function isomorphism. First, the imperfect and the pluperfect are analysed as encoding a *remote* tense, which subsumes both the temporal and the conditional interpretations of these TAM sets in the traditional analysis. Consequently, there is no need to recognise a specific conditional mood. Second, the preterite and the pluperfect are considered to encode the perfective aspect, while all the other TAM values are *non-perfective* by default. Table 13.1 shows the semantic analysis of indicative TAM values as used in the following discussion.

Table 13.1. Indicative TAM values

Value	Tense	Aspect
Present	non-remote	non-perfective
Future	future (non-remote)	non-perfective
Imperfect	remote	non-perfective
Preterite	non-remote	perfective
Pluperfect	remote	perfective

Table 13.2. Mismatching TAM values in lexical verbs and in the copula

Function	Lexical verbs	Copula
Indicative past perfective	preterite (*kerďal*)	past (*ssalahi*)
Indicative past non-perfective	imperfect (*keresahi*)	
Conditional real/potential		conditional (*ovesahi*)

The present, the future, the pluperfect, and the imperative in the copula have the same functions as the corresponding TAM values in lexical verbs. Apart from the present subjunctive (e.g. *ovas* '[that] we are'), there is also a *past subjunctive* (e.g. *uliljam* '[that] we were'). The greatest mismatch between the TAM paradigms of the lexical verbs and the copula is charted in Table 13.2 (examples from Šóka Rumungro). While lexical verbs differentiate aspect in the indicative past and, at the same time, conflate the real/potential conditional with the non-perfective past, the copula possesses a single, aspect-indifferent, past value (the past) and a distinct real/potential *conditional* value.

Forms of the remote tense tend to be more complex, and are more likely to be eroded, than forms of the non-remote tenses. However, there is no clear complexity asymmetry between the (remote) imperfect and the (non-remote) future. Among non-remote values, the future tends to be more complex than the present. In the copula, past forms are more complex than present forms. Forms of the remote tense are less differentiated than forms of the non-remote values. Present verb forms are more likely to be borrowed than forms of other tenses. As for aspect, the perfective is the more complex value. Both directions of aspect extension are attested, and there are conflicting differentiation asymmetries. As for mood, the imperative is the least complex and the least differentiated value, while the indicative is the most complex and the most differentiated value. There are conflicting extension asymmetries in mood. The subjunctive shows intermediate complexity and differentiation, and a greater extracategorial distribution than the other moods. Aktionsart modifications are more complex than unmodified verbs, and markers of aktionsart modification may be borrowed and show certain cross-dialectal diversity (as against 'neutral' aktionsart).

13.1. Complexity

The imperative is the only zero coded TAM term: with most verbs the second-person singular imperative coincides with the inflectional stem. There

are also several instances of relative zero coding. The following complexity hierarchy holds among the non-perfective values: imperative < subjunctive < present < future, imperfect. The mutual position of the future and the imperfect is ambiguous. As for perfective values, the pluperfect is more complex than the preterite. The perfective is more complex than the non-perfective in that it is marked by overt perfective markers. There are no specific aspect markers in non-perfective forms.[1]

Within the non-remote non-perfective domain, the subjunctive tends to be the least complex value, and the future (indicative) tends to be the most complex value, while the present (indicative) assumes an intermediate position. There are four types of structures in the non-remote non-perfective domain. First, the so-called short forms, which do not show any overt TAM marking (e.g. *kerav*, 1SG of 'do'). Second, the so-called long forms, which contain the word-final suffix *-a* (e.g. *kerav-a*). Both the short forms and the long forms have been inherited from Early Romani. Third, Core Sinti dialects have developed secondary short forms from the long forms through erosion of the suffix *-a*. The secondary short forms constitute a distinct set, which is different from both the primary short forms and the long forms: e.g. secondary short *keraw* (< **kerav-a*), primary short *kerap* (< **kerav*), and long *kerava* in Austrian Sinti. Finally, there are two sorts of analytic future forms. In dialects spoken in the Balkans, both Balkan and Vlax, they consist of a proclitic future particle plus the short or, rarely, the long form. The future particle has, in convergence with the Balkan languages, developed through grammaticalisation of the verb 'want' (e.g. *kam*, *kan*, *ka* < **kam-*, or *ma* < **mang-*). In Russian Romani and Prizren Arli, analytic future forms consist of an auxiliary verb (*l-* 'take' or *av-* 'come' in the former, third-person copula in the latter) and a short finite form of the lexical verb introduced by a non-factual complementiser (e.g. Russian Romani *lav te kerav* 'I will do' < *'I take that I do', Prizren *si te kerav* 'I will do' < *'I must do' < *'it is that I do').

Dialects differ in whether they possess any analytic future constructions, and in the functional distribution of the short and the long synthetic forms. The short forms may function as subjunctive or subjunctive-present forms, and the long forms may function as future, present, or present-future (i.e. non-past indicative). Table 13.3 shows seven major types of marking in the non-remote non-perfective domain.

Type A1, found in Finnish Romani, Piedmontese Sinti, and Taikon Kalderaš, preserves the Early Romani distribution, where the long forms encode the non-past indicative, as against the short subjunctive. Type A2, found in Core Sinti dialects, shows an identical division; in the non-past indicative, however, the

Table 13.3. Non-remote non-perfective forms

	Subjunctive	Present	Future
Type A1	short	long	
Type A2	short	long ~ secondary short	
Type B	short	long	
Type C1	short		long
Type C2	short		analytic
Type D	short	long	analytic
Type E	short	long	analytic

long forms alternate with secondary short forms.[2] Type B is represented by Welsh Romani, Latvian Romani, and older Rumungro. Here, the subjunctive is encoded by the short forms, the future by the long forms, and both forms may be used in the present. In Type C1, the long forms are specialised for the future, while the short forms encode the subjunctive-present. This type is found in a continuous area in Central-East Europe: in Slovene Romani, the Central dialects, the western part of the Northeastern dialects (Polish, Lithuanian, and Podolie Romani), Lovari, and Ukrainian Vlax. Type C2 shows an identical division, with the subjunctive-present encoded by the short form and the future by an analytic construction; long forms are specialised for modal uses or they have been lost (e.g. in Priština Gurbet). This type is typical of South Vlax dialects (e.g. Xoraxane, Dasikano, Priština Gurbet, Ajia Varvara, Varna Kalajdži, Rešitare, and Kalburdžu), and is also attested in Sofia Erli, Gadžikano, and Bunkuleš Kalderaš. Type D, found in numerous Balkan dialects (e.g. Arli of Prizren and Florina, Sepečides, Rumelian and Iranian Romani, Yerli, Bugurdži of Varna and Kosovo, and Kaspičan), exhibits a tripartite division between the short subjunctive, the long present, and the analytic future. Type E is similar, except that the present may be encoded by both the long and the short forms. This type is attested in Russian Romani and some dialects of the Balkans (e.g. Arli of Prilep, Skopje and Gilan, Crimean Romani, Malokonare, Nange, Muzikanta, and Vălči Dol).

In all dialects, the subjunctive is encoded by the short form. In other words, it has zero TAM coding. The future shows the greatest complexity: it is either encoded by the long form or by an analytic construction. The present assumes

an intermediate position: it may go together with the future (Type A) or with the subjunctive (Type C), it may be split between the two (Type B), or it may possess a form of its own that is of intermediate complexity (Types D and E). The cross-dialectal evidence results in the complexity hiearchy: future > present > subjunctive.

Remote categories tend to be more complex than non-remote categories: the imperfect (i.e. the remote non-perfective) tends to be more complex that the other non-perfective categories, and the pluperfect (i.e. the remote perfective) is more complex than the other perfective categories. The pluperfect is mostly derived from the corresponding preterite forms by addition of the remoteness marker (e.g. Early Romani **kerdjom-asi* 'I had/would have done' < **kerdjom* 'I did'), although in a few dialects the pluperfect is homonymous with the preterite in some person–number combinations (see Chapters 6 and 7). And the imperfect is derived from the corresponding short non-remote (i.e. subjunctive or subjunctive-present) forms by addition of the remoteness marker (e.g. Early Romani **kerav-asi* 'I was doing, I would do' < **kerav* '[that] I do'). In other words, the preterite is zero coded with regard to the pluperfect, and the subjunctive or subjunctive-present is zero coded with regard to the imperfect.

The relation between the long non-remote forms and the imperfect is more ambiguous. In Early Romani and in most dialects, they both contain a single overt TAM marker, and so they show the same degree of morphological complexity. Nevertheless, the imperfect tends to be more complex phonologically. In most dialects, the reflexes of the remoteness suffix **-asi* (*-ahi*, *-as*, *-ys*) are longer than those of the suffix **-a* of the long forms, in terms of number of phonemes or even syllables.[3] Only in a few dialects is there no such asymmetry due to erosion in the remoteness suffix. In modern Core Sinti and in Slovene Romani, both suffixes are monophonemic: cf. the third-person singular forms in Manuš *kerel-a* vs. *kerel-s* (*-s* < **-es* < **-as* < **-asi*), and in Slovene Romani *kerel-a* vs. *kerel-e* (possibly *-e* < **-ai* < **-ahi* < **-asi*). In Kalburdžu and Xoraxane, the remoteness suffix has lost the final /s/ due to a regular phonological process, and so the imperfect has been conflated with the original forms in *-a*. Synchronically, we have a single set of forms which encode a wide range of TAM functions (e.g. Kalburdžu *keras-a* '[if] we do; we were doing/ used to do, we would do').

A few dialects (some Arli varieties, Abruzzian Romani, and possibly Slovene Romani) present additional evidence for a greater complexity of the imperfect with regard to the long non-remote forms. These dialects have innovated remoteness marking by grammaticalisation of the third-person past cop-

ula form and its cliticisation to the present (long) forms (e.g. Prizren Arli *kerela-sine* 's/he was doing' < *'s/he does-was').[4] The cliticised copula tends to further reduce in shape (e.g. Skopje Arli *kerela-ine*), turning into a suffix and even fusing with person–number suffixes in Abruzzian Romani (cf. first person *kera-snə*, and second and third person *kere-snə*). It has been suggested that the copula may also be the source of the Slovene Romani remoteness marker *-e* (Cech and Heinschink 2001, see above for an alternative scenario). At an initial stage of grammaticalisation, the new imperfect is more complex than the long forms in terms of the number of morphemes it contains.

In those dialects where the future is encoded by the long forms (Types A, B, and C1 in Table 13.3), the imperfect is more complex than, or at least as complex as, the future. However, in dialects that possess an analytic future (Types C2, D, and E in Table 13.3) and, at the same time, the original synthetic remoteness marker, it is the future that is more complex than the imperfect. In Arli of Skopje, Gilan, and Prizren, both categories are encoded analytically, and there is no obvious asymmetry in complexity.

In the copula, the present tends to be less complex than the past in that the former is more likely to select the 'weaker' root *h-* in the third person (as against the 'stronger' root *s-*; see also Chapter 7) than the latter. This generalisation may be formulated as a unilateral implication: if a dialect has the root *h-* in the past, it will also have it in the present. Also, the third-person past forms are more likely to contain the participial suffix *-in-* than the third-person present forms, as found in some Southern Central dialects and numerous dialects of the Balkans (e.g. Erli *s-in-e* 's/he was, they were' but *s-i* 's/he is, they are'), although in a few dialects the opposite holds (e.g. Finnish Romani *h-in* 's/he is, they are' but *s-as* 's/he was, they were'). In other respects, the complexity asymmetries in the copula parallel those in lexical verbs, as described above.

Aktionsart modifications are frequently expressed by indigenous or borrowed modification adverbs (e.g. Rumungro *sīt dža-* 'dissolve, lit. go apart', with *sīt* from Hungarian), borrowed prefixes (e.g. East Slovak Romani *roz-dža-* 'dissolve', with *roz-* from Slovak), or through indigenous or borrowed auxiliary matrix verbs (e.g. Rumungro *sokin- te keren* 'do habitually', lit. 'be accustomed to do', calquing Hungarian and borrowing the matrix verb). The iterative, which has become productive in the Central dialects, is the least complex aktionsart in that it is marked by an indigenous suffix (a grammaticalisation of the verb *ker-* 'do') that may show a certain degree of fusion with the stem. Any aktionsart modification is more complex than its unmodified base verb.

13.2. Erosion

Several dialects (e.g. Welsh Romani, Finnish Romani, some Sinti varieties, Latvian Romani, the South Central dialects, and Lovari) contract the first-person singular non-perfective inflection -(*a*)*v*- with a following TAM marker. In most dialects, the contraction affects both the long and the imperfect forms (e.g. Slovak Lovari *ker-ow* 'I will do' < **ker-av-a*, and *ker-ows* 'I was doing' < **ker-av-as*). In Welsh Romani, however, the imperfect is contracted (e.g. *ker-ās*), while the long (present-future) forms remain uncontracted (e.g. *ker-ava*). A similar situation holds in Prilep Arli, where the imperfect but not the long (present) forms are optionally contracted, and in Manuš, where the imperfect is contracted obligatorily, the long (present-future) forms only optionally. Thus, erosion appears to affect the imperfect (the remote non-perfective) more than the non-remote non-perfective forms.

13.3. Differentiation

The criterion of differentiation renders a clear tense asymmetry and ambiguous aspect asymmetries. As for tense, non-remote forms are consistently more differentiated than remote forms, except with regard to inflectional classification, where there is no asymmetry; and, in the copula, present forms are more differentiated than past forms. As for aspect, perfective forms are more differentiated in evidentiality and inflectional classification, and less differentiated in number. The perfective is also more likely to be more differentiated in person, although a few dialects exhibit the opposite asymmetry. There are no differentiation asymmetries among the non-remote non-perfective categories (the subjunctive, the present, and the future). The imperative stands out among the non-perfective forms in that it contributes its own dimension of classification, and in that it is defective (i.e. less differentiated) in person. There are no aktionsart asymmetries in terms of differentiation of inflectional categories.

Differentiation of **person** tends to be, with some exceptions, greater in perfective forms than in non-perfective forms (irrespective of tense). In Early Romani and in many dialects, person homonymy (viz. homonymy of the second and third persons in the plural, see Chapter 7) is restricted to non-perfective forms, and so perfective forms commonly show greater person differentiation than non-perfective forms. There are no dialects with no homonymy in the non-perfective sets. However, numerous dialects (e.g. most Northeastern dialects, Sinti, Yerli, and Banat Kalderaš) have copied the inherited non-perfective

person homonymy to the perfective through various person extensions (see Chapter 7), and so now there is no asymmetry between the non-perfective and the perfective with regard to person differentiation. Subsequent developments in East Ukrainian and Podolie Romani have resulted in a greater extent of person homonymy in the perfective than in the non-perfective: while in non-perfective forms only the second- and third-person plural are homonymous, there is complete person neutralisation in the plural of perfective forms (see Chapter 7). Person also tends to be more differentiated in non-remote forms than in remote forms (irrespective of aspect). In modern German and Austrian Sinti (see Table 6.8 in Chapter 6), the second and the third persons are generally homonymous in remote forms, but only in the plural of non-remote forms.

In two dialects, asymmetry with regard to person differentiation must be stated in terms of a combination of aspect and tense. In Hungarian Sinti, the second- and third-person plural are homonymous in all finite forms, while the pluperfect shows also a homonymy between the second- and third-person singular. In Roman, the preterite stands out in being the only finite set that does not show any person homonymy (Table 13.4).

Whereas the greatest differentiation of the Roman preterite (i.e. the non-remote perfective) corroborates the asymmetries found independently in the categories of aspect (perfective > non-perfective) and tense (non-remote > remote), the least differentiation of the pluperfect (i.e. the remote perfective) in Hungarian Sinti represents another exception to the tendency of the perfective to be more differentiated in person than the non-perfective.

Differentiation of **number** is greater in non-perfective forms than in perfective forms, and within the latter, it is more likely to be greater in non-remote (preterite) forms than in remote (pluperfect) forms. Number homonymy in some persons either occurs in equal extent in both perfective forms (e.g. in Sofia Erli, Rumelian Romani, Polish and Hungarian Lovari, Taikon Kalderaš,

Table 13.4. Verb inflections in Roman

	1SG	2SG	3SG	1PL	2PL	3PL
PRES-SUBJ	*-av*	*-es*	*-el*	*-as*	*-en*	
FUT	*-a*	*-eh-a*	*-l-a*	*-ah-a*	*-n-a*	
IMPF	*-a-hi*	*-eh-ahi*	*-l-ahi*	*-ah-ahi*	*-n-ahi*	
PRET	*-om*	*-al*	*-a*	*-am*	*-an*	*-e*
PLPF	*-om-ahi*	*-al-ahi*	*-a-hi*	*-am-ahi*	*-an-ahi*	

and Vălči Dol), or it is restricted to the remote perfective (in Zemplín Slovak Romani), or the remote perfective shows greater extent of number homonymy than the non-remote perfective (in Austrian Lovari and Welsh Romani). As for asymmetries within the remote perfective domain, Bougešťi exhibits number homonymy in the irrealis conditional form, but not in the pluperfect proper. (See Chapter 7 for details.) There is no number homonymy in non-perfective forms in any dialect.

Perfective forms show more differentiation than non-perfective forms in that, in some dialects, some verbs possess two forms in the perfective third-person singular: a finite form and an active participle. The distinction may be employed to encode **evidentiality**. In a few dialects (e.g. Taikon Kalderaš and older Prilep Arli), the distinction between the finite and the participial perfective is only retained in the preterite (e.g. Prilep *gelo* 'he went' and *geli* 'she went' vs. *gelas* 's/he went'), while there are only finite forms in the pluperfect (e.g. *gelasas* 's/he had gone').[5] Thus the distinction is neutralised in the remote perfective, and so the non-remote tense shows more differentiation than the remote tense.

Inflectional **classification** in verbs is generally carried out by the shape of non-perfective person–number suffixes, the shape of (second-person singular) imperative suffixes, and the shape of perfective suffixes. As for non-perfective person–number suffixes, there are three to five non-perfective classes per dialect. Early Romani probably possessed three non-perfective classes: one for oikoclitic consonant stems (e.g. 3SG *-el*), one for oikoclitic vocalic stems (e.g. 3SG *-l*), and one for xenoclitic verbs (3SG *-i*). The difference between the two oikoclitic classes could be accounted for by morphophonological rules, viz. deletion of the initial vowel of the non-perfective person–number suffix after the final vowel of a vocalic stem (e.g. [*dža-*][*-el*] > *dža-l* 's/he goes'). The xenoclitic class has been lost in many dialects where borrowed verbs now inflect as oikoclitics (see Chapter 23). Most dialects have created further non-perfective classes through contractions in middle verbs of middle suffix(es) and the non-perfective person–number markers.[6] For example, there are four classes in Ajia Varvara (Table 13.5): two in active verbs (the consonantal stems and the vocalic stems), and two in middle verbs, viz., the class of the (Early Romani) middles in **-jov-* and the class of the (Vlax) middles in **-áv-*.

Due to the middle contractions, the variants of the person–number inflections are no longer accountable for by morphophonological rules. Person–number inflections of some sets (viz. those of the active consonant stems and those of the middle classes) are now best analysed as containing a classificatory segment (e.g. Ajia Varvara *-a-* ~ *-e-*, *-a-* ~ *-o-*, and *-ia-* ~ *-o-*) and a per-

Table 13.5. Imperfective verb classes in Ajia Varvara

	1SG	2SG	3SG	1PL	2PL	3PL
Active (V-stems)	*-v*	*-s*	*-l*	*-s*	*-n*	*-n*
Active (C-stems)	*-a-v*	*-e-s*	*-e-l*	*-a-s*	*-e-n*	*-e-n*
Middle (*-*áv*-)	*-a-v*	*-o-s*	*-o-l*	*-a-s*	*-o-n*	*-o-n*
Middle (*-*jov*-)	*-ia-v*	*-o-s*	*-o-l*	*-ia-s*	*-o-n*	*-o-n*

son–number suffix proper. Person–number inflections of active vowel stems consist of the person–number suffixes alone. The morphological status of the classificatory segments is evidenced by various morphological extensions (e.g. by the extension of *-o-* from the middles in *-*jov*- into the middles in *-*áv*-).

Disregarding individual irregular verbs, there are usually no more than three imperative classes. In Early Romani, most verbs had no overt (second-person singular) imperative marker, while monoconsonantal stems (viz. *d-* 'give' and *l-* 'take') and their derivatives employed the suffix *-e*. The distribution of this suffix has been narrowed in some dialects (e.g. Slovene Romani, and some South Central and Arli varieties), and extended in others. In some Balkan dialects (e.g. Arli of Prilep and Florina, Sepečides, and many Bulgarian dialects) the suffix spread to the imperative of adapted loan verbs, in some other dialects it spread to verbs of certain derivational structures (e.g. to the middles in Xoraxane) etc. Finally, most dialects possess a few verbs with a stem in /j/, whose imperative ends in /i/ (e.g. *uštj-* 'get up, rise' > *ušti*, *urj-* 'dress' > *uri*). In Early Romani and in many dialects, the alternation /j ~ i/ is accountable for by morphophonological rules, and so there is no overt imperative marking. However, in some dialects, sound changes in these verbs have resulted in a reanalysis of an imperative suffix *-i*, and hence in creation of a new imperative class (e.g. North Central *uštʲ-* > *uštʲ-i*, *ur-* > *ur-i*).

The number of perfective classes is greater. Early Romani possessed at least seven perfective markers (*-d-*, *-t-*, *-n-*, *-in-*, *-l-*, *-il-*, *-ist-*), plus a few individual verbs that showed, in addition, an irregular stem alternation. Dialects show various re-assignments of the perfective markers among different classes, and frequently also reinforcement through concatenation of several perfective markers within a verb form (e.g. *-d-l-*, *-n-il-*, *-in-d-*). Individual dialects usually retain or even enlarge the wealth of Early Romani perfective classes. Only in modern Sinti dialects is there a tendency towards reduction of perfective classification due to the gradual loss of the perfective suffix.

The imperative is the least differentiated finite category in that it is *defective* in terms of person distinctions: there are only second person imperative

forms. Hortative constructions of the other persons involve subjunctive forms with or without a specific hortative particle (e.g. *mi* grammaticalised from the verb *mek-* 'let, leave'), depending on dialect.

Indicative forms of the copula show greater differentiation than non-indicative forms and, within the former, present forms show greater differentiation than past forms, in that the indicative and the present are more likely to co-occur with clitic subject pronouns than the non-indicative and the past, respectively. Following the development of a new set of third-person nominative pronouns in *o-*, the original forms in *l-* (which became cliticised and even suffixed in some dialects) have been retreating in Romani. The majority of dialects have either lost the subject clitics completely or restricted their distribution to the copula and/or non-verbal predications (see Chapter 21 for further details). Some dialects do not impose any TAM restrictions on the distribution of the subject clitics with the copula (e.g. Sinti and Roman); in others (e.g. Šóka and Nógrád Rumungro, and Prizren Arli) the clitics can only co-occur with the present and the past, i.e. with forms with the indicative root *s-* or *h-* (e.g. Šóka Rumungro *hi-lo* 'he is' and *sāhi-lo* 'he was' but not **ovla-lo* 'he will be'); and in most dialects that retain the clitics with the copula at all (e.g. Finnish and Lithuanian Romani, most Central dialects, Slovene Romani, Prilep Arli, Austrian Lovari, Rakarengo, and Kalderaš), they are restricted to the present forms (e.g. East Slovak Romani *hin-o* 'he is' but not **sas-o* 'he was').

13.4. Extension

Aspect extension is attested in both directions. An extension within the category of tense is difficult to interpret (see below). There are conflicting directions of extension in the category of mood: subjunctive forms may extend into the indicative (in verbs) but also vice versa (in verbs and possibly in negators); and indicative forms may extend into the imperative but, according to one diachronic scenario, also vice versa (in negators).

In Early Romani and most dialects, there are a few verbs that show an irregularity in the formation of their perfective stems. While most verbs derive their perfective stem by suffixation of a perfective marker to the non-perfective stem (e.g. NPFV *ker-* > PFV *ker-d-* 'do'), some irregular verbs in addition undergo an irregular stem alternation (cf. *rov-* > *ru-n-* 'weep', *sov-* > *su-t-* 'sleep', *mer-* > *mu-l-* 'die', *per-* > *pe-l-* 'fall') or even suppletion (cf. *dža-* > *ge-l-* 'go'). Some dialects, especially Welsh Romani and the modern Northwestern dialects, have abandoned this irregularity by extending, optionally or

obligatorily, the non-perfective stem to the perfective forms in some or all of the irregular verbs. The non-perfective stem of the verb *rov-* 'weep' has extended to the perfective in Welsh Romani (e.g. *rov-d-om* 'I weeped'), Finnish Romani (e.g. *rouv-id-om*), Austrian Sinti (e.g. *rov-um*), some Northeastern dialects (e.g. *ro*(*v*)*-dž-om*), Kosovo Bugurdži (e.g. *rov-dz-om* or *rov-om*), and some South Vlax dialects (e.g. Priština Gurbet and Vălči Dol *rov-d-em* or Ajia Varvara *rov-l-em*). In some Balkan and South Vlax dialects, the non-perfective stem influenced the perfective stem in the quality of its vowel (e.g. Sofia Erli *ro-nj-om*, Nange, Varna and Kaspičan *ro-j-om*, and Kalburdžu *ro-em*). The stem of the verb *sov-* 'sleep' has been regularised in Welsh Romani (e.g. *sov-d-om* 'I slept'), Finnish Romani (e.g. *souv-id-om* or *sov-j-om*), and Austrian Sinti (e.g. *so-d-um*). Regular formations of the verbs *mer-* 'die' and *per-* 'fall' occur in Welsh Romani (e.g. *mer-d-om* 'I died', and *peř-d-om* 'I fell'), some modern varieties of Finnish Romani (e.g. *mer-t-om* and *per-t-om*), Austrian Sinti (e.g. *mer-d-um* and *per-d-um*), and Kaspičan (e.g. *mer-ij-om* and *per-ij-om*). Finally, the non-perfective stem of the suppletive verb *džа-* 'go' has extended to the perfective in some Sinti dialects (e.g. Austrian Sinti *dž-um*, and Manuš *dž-j-om*, both from **dža-j-om*).[7]

In Rumungro, there is a grammatically conditioned alternation of the non-perfective second-person and the first-person suffixes before tense suffixes. While the future suffix *-a* triggers the alternation */s* > *h/*, resulting from phonological erosion, the remoteness suffix *-ahi* does not (e.g. *ker-ah-a* 'we will do' vs. *ker-as-ahi* 'we were doing'). Since earlier remoteness forms, in all likelihood, did involve the alternation, i.e. had undergone the erosion (as in the closely related Vendic dialects, e.g. *ker-ah-ahi*), the non-alternating remoteness forms must have resulted from a morphological extension of the */s/* from corresponding present-subjunctive forms (e.g. *keras* 'we do').[8] In a few Rumungro varieties, the extension is further proceeding to the future forms as well: the youngest speakers of Šóka Rumungro now prefer the innovative non-alternating forms (e.g. *ker-as-a* 'we will do') over the alternating forms of the older speakers. The extension thus proceeds from a non-remote form (the present-subjunctive) to a remote form in all relevant dialects, but then again to a non-remote form (the future) in a subgroup of these dialects. This sort of extension is difficult to interpret in terms of the category of tense.

Cross-dialectal evidence suggests that, in Early Romani, adaptation of borrowed active verbs involved the Greek suffixes *-Vz-* or *-Vn-* in the non-perfective, and the Greek aorist suffix *-Vs-* in the perfective (cf. Matras 2002: 128–133). In the Northeastern, the Central and most Balkan dialects, and also in some Lovari varieties, the non-perfective adaptation markers have extend-

ed to the perfective (e.g. East Slovak Romani *us-in-d-e* 'they swam', Kosovo Bugurdži *piš-iz-d-e* 'they wrote'), while in most Vlax dialects and partly also in Welsh Romani, the perfective adaptation marker has extended to the non-perfective (e.g. Varna Kalajdži *vorb-is-ar-en* 'they speak').

Mood extensions are well represented in negators. Early Romani possessed two distinct verb negators: the indicative *na* and the imperative *ma* (the issue of which of them was used in the subjunctive will be addressed below). Numerous dialects have developed new indicative negators. The negator *či* in the North Vlax dialects as well as in some adjacent South Vlax dialects (e.g. Gurbet of Srem and Bačka) is a result of grammaticalisation of the Early Romani indefinite **či* 'something, nothing' (cf. Elšík 2000c). A similar grammaticalisation path gave rise to the negator *kek* in Welsh Romani, which still retains its original indefinite function as a negative determiner 'no, none'. Most South Vlax dialects and a few North Vlax varieties of Romania possess the negator *ni* (e.g. Dasikano, Kumanovo Gurbet, Vidin Cocomanya, Lom, Rešitare, Ajia Varvara, and Rakarengo) or, in the east, *in* (e.g. Kalburdžu, Vălči Dol, and Varna Kalajdži), whose origin is not clear. Finally, there is also borrowing of indicative negators in modern Sinti dialects and Finnish Romani (see Section 10.5). Table 13.6 shows four types of patterns in the indicative and imperative negators.

Type A engenders the Early Romani distribution of negators, which is retained in older Sinti varieties, most Central dialects, Slovene and Abruzzian Romani, and almost all the Balkan dialects. Some South Vlax dialects (e.g. Kumanovo Gurbet, Ajia Varvara, Rešitare, Kalburdžu, Vălči Dol, and Varna Kalajdži) and modern Sinti represent Type B, where the indigenous indicative negator *na* has been replaced by an innovative form. Welsh and Finnish Romani are transitional between Types A and B, in that they retain the indigenous indicative negator but also possess an innovative variant. Type C shows an extension of the indigenous indicative negator to the imperative as well, so that, synchronically, there is a single negator that is not sensitive to the

Table 13.6. Indicative and imperative negators

	Indicative	Imperative
Type A	*na*	*ma*
Type B	innovative	*ma*
Type C	*na*	*na*
Type D	innovative	*na*

mood of the verb. The extension has come about independently in various dialects (e.g. in the Northeastern dialects, in Šóka Rumungro, in Cerhari, and in Kaspičan and Gadžikano). An identical extension must also be assumed for dialects of Type D, viz. the North Vlax and some South Vlax dialects (e.g. the northern Gurbet-like varieties, Lom, and Cocomanya). Here, the original indicative negator must have first extended to the imperative before it was replaced by an innovative negator in the indicative value.

In Table 13.6 and in the above discussion, we have disregarded negators of verbs in the subjunctive mood. Consider now Table 13.7, where subjunctive negators are added. We disregard the Type C pattern here, since it has a single mood-indifferent negator, and add another pattern (Type E).

In most dialects, the subjunctive negator is homonymous with either the indicative or the imperative negator: the subjunctive assumes an intermediate position between the two other mood values. Only in a few South Vlax dialects (e.g. Kalburdžu, Vălči Dol, and Ajia Varvara),[9] is there a distinct subjunctive negator, in addition to an indicative one and an imperative one (Type E). Each of the other types may be divided into a subtype with a subjunctive-indicative negator (A1, B1, D1), and a subtype with a subjunctive-imperative (or non-indicative) negator (A2, B2, D2). Type A2 is found in Slovene Romani, Prizren Arli, and Sepečides, and Type D2 in the North Vlax dialects and in some South Vlax dialects (e.g. northern Gurbet-like varieties and Cocomanya); Type B2 is unattested.

Now, which direction of mood extensions one recognises depends on how one reconstructs the original distribution of negators. Assuming that Early Romani was of Type A1 (e.g. Boretzky 2003: 56), there must have been an

Table 13.7. Indicative, subjunctive, and imperative negators

	Indicative	Subjunctive	Imperative
Type A1	*na*		*ma*
Type A2	*na*	*ma*	
Type B1	innovative		*ma*
Type B2	innovative	*ma*	
Type D1	innovative		*na*
Type D2	innovative	*na*	
Type E	innovative	*na*	*ma*

extension of the imperative negator *ma* to the subjunctive in Type A2 (and the unattested B2). Assuming, on the other hand, that Early Romani was of Type A2, there must have been an extension of the indicative negator *na* to the subjunctive in Types A1, B1, and E. Types D1 and D2 as well as C in any case require an extension of an indicative or an indicative-subjunctive negator to the imperative. We may conclude that the indicative extends to the imperative (both scenarios), possibly via the subjunctive (the second scenario); and that the imperative possibly extends to the subjunctive (the first scenario).

In Hungarian Sinti, the long (indicative present-future) verb forms take over the subjunctive in the plural (see Table 6.9 in Chapter 6 for the paradigm). This confirms the possibility of an indicative-to-subjunctive extension. On the other hand, the short (subjunctive) forms commonly extend to the indicative present (see Section 13.1).

13.5. Extracategorial distribution

The subjunctive shows a greater extracategorial distribution than the other TAM values in that dialects that have created the so-called new infinitive employ subjunctive forms in this function (see Chapter 7).

13.6. Borrowing

Present verb forms are more likely to be borrowed than forms of other tenses. A number of dialects of Bulgaria borrow the impersonal necessitative auxiliary *trjabva* 'is necessary' from Bulgarian. Some of these dialects form the past by means of indigenous morphology (e.g. Muzikanta *trjabv-as* 'was necessary'), while others borrow the past form from Bulgarian (e.g. Malokonare, Nange, Šumen Drindari, and Rešitare *trjabvaše*). In Kosovo Bugurdži, a necessitative auxiliary is borrowed from Serbian in its person-inflected present forms (e.g. *mora-m* 'I must'), while the past forms are formed by means of indigenous morphology (e.g. *mora-nj-om* 'I had to'). In various dialects, the Greek-derived third-person singular marker *-i* is used in present forms of some verbs, while future and imperfect forms contain the indigenous third-person singular marker *-(e)l* (e.g. Slovene Romani *ker-i* 's/he does', but *ker-el-a* 's/he will do' and *ker-el-e* 's/he was doing'). For borrowing of markers of aktionsart modification see Section 13.1.

Chapter 14
Modality

Modality is a dependency relation which constrains the truth value of a proposition. The overall feature of the category is the depiction of non-real (non-factual) states of affairs, but there are several sub-categories. Strict dependency modality creates a bond between the outcome of the predication depicted by the modal modifier and the realisation of the main event. Among such modal modifiers, we might draw distinctions such as mental or physical states (volition, ability, fear, need, wish) on the one hand and actions (attempts, beginnings, and so on) on the other, as well as between internal and external forces ('want' vs. 'must'). Conditional modality or *conditionality* is the dependency of a proposition on a conditional predication. Here too there are sub-types (conditionality values), such as real, potential, or unreal (irrealis). Further types of modality include conditional complements ('whether'), optatives and imperatives (statements of volition or manipulation), and evidentials (the latter constrain the truth value to what is inferrable from a situation).

Perhaps the most outstanding feature of modality in Romani, not unlike other languages of the Balkans (cf. Friedman 1985), is its structural exposition through the modal conjunction *te*. This connector introduces optative predications, modal complements, direct and indirect conditional clauses, purpose clauses, as well as different types of adverbial subordinations, such as different-subject simultaneity ('I heard them talking') and negative-circumstance ('he drank without spilling'). The connector tends to occupy a position immediately preceding the verb, and is followed by the subjunctive, or, in some dialects, by a generalised subjunctive form, which latter in effect is a person-neutral as well as tense-neutral form which has been referred to as the 'new infinitive' (Boretzky 1996b; cf. also Matras 2002: 161–162).

Dependent modality is characterised by low complexity, while conditionality is arguably more complex than its indicative counterpart structures. Within expressions of modality, there is a split between those expressing ability and necessity, and those expressing volition or actions (e.g. 'begin' or 'try') as well as other mental states ('fear', 'dare'). The former may be less complex in their relation to their complements and less differentiated in their inflectional potential than the latter. On the scale of borrowability, necessity ranks

highest, followed by ability. While the marking of dependent modality itself is extremely stable in the language, conditionality often shows borrowings.

14.1. Complexity

Clause linking through dependent modality tends to be less complex than the linking of independent indicative clauses, provided that the subjects of the two linked clauses are identical. Dependent modality triggers the subjunctive in the complement clause, which in many dialects is less morphologically complex than the indicative (see Chapter 13). In dialects that allow absence of agreement on the complement verb ('new infinitive'), the complement verb is also less differentiated in comparison with indicative, finite verbs of main clauses. To this we can add the omission of the identical subject in the complement clause (equi deletion). Both features may be seen to be conditioned not by modality itself, but by the linking of dependent events, and so in effect they are properties related to the *discreteness* of the events that are being linked (see Chapter 12).

Conditionality on the other hand tends to be more complex than the corresponding non-modal (indicative) forms in the composition of the verb, the obligatory presence of a conditional conjunction in the protasis, the obligatory presence of a following apodosis, and often the presence of a (borrowed) conditional particle in either part of the construction. Table 14.1 shows patterns of distribution of tenses among types of conditional constructions. Most dialects differentiate between realis, potential, and irrealis conditionality. Differences are found in the distribution of individual tense forms, in the presence or absence of a specific conditional form of the tense (marked CTENSE in the Table 14.1), and especially in the extent to which there are matching tense forms in the protasis and apodosis of the conditional clause.

In Type A, both parts show the same tense. This pattern is found among the central European dialects (North Vlax, Central, Sinti, Slovene Romani, and Polish Romani). While there is variation among the choice of tense pairs in the realis, only Polish Romani shows perfective forms in the potential. Types B–E are characterised by the presence of a conditional mood, marked by the particle *ka*(*n*) in combination with tense inflection. These types are found exclusively in the Balkans. Type B is common among the dialects of Bulgaria. It is characterised, like Type A, largely by matching tenses in the protasis and apodosis, with the exception of some combinations in the realis. But unlike Type A it shows consistent mood mis-match, with a conditional mood marker in the

Table 14.1. Distribution of tenses among types of conditional clauses

	Realis	Potential	Irrealis
Type A	PRES–PRES FUT–FUT	IMPF–IMPF PERF–PERF	PLPF–PLPF
Type B	SUBJ–CSUBJ FUT–CSUBJ PRES–CSUBJ	IMPF–CIMPF	PLPF–CPLPF
Type C	PRES–PRES CSUBJ– CSUBJ	IMPF–CIMPF	PLPF–CPLPF
Type D	FUT–CSUBJ	SUBJ–CIMPF	IMPF–CPLPF
Type E	PRES–PRES	PERF–CSUBJ	PERF–CIMPF PLPF–CIMPF
Type F	PRES–PRES, PRES–IMPF	SUBJ-IMPF	PLPF–PLPF
Type G	PRES–PRES, PRES–IMPF		IMPF–IMPF PRET–PRET

apodosis. Type C (Epiros) is essentially similar, but shows no mood opposition between protasis and apodosis in the realis. By contrast to the previous types, Type D (Crimean Romani) shows both tense and mood mis-match in protasis and apodosis. Type E (Karditsa Arli) is again similar, but has no tense or mood opposition in the realis. Type F (Lithuanian Romani) lacks mood oppositions, but makes use of tense oppositions in the potential, and optionally in the realis. Type G finally (Finnish Romani), shows no realis:potential opposition, and matching of the tenses in the irrealis.

Non-matching tense forms are somewhat more likely to appear in the realis, followed by the potential, and finally the irrealis. Apart from that, the data show no obvious asymmetries between the individual conditionality types. Within the conditional clause, however, the apodosis tends to be more complex than the protasis, taking into account both the frequent presence of the conditional mood marker in the apodosis, and the frequent presence of more complex tense forms in the apodosis.

Another complexity scale involves the type of connecting device that links the modal (verb or uninflected particle) with its complement (Table 14.2). Some impersonal modals such as *šaj/ašti/naj* ‘can’, *našti* ‘cannot’, tend not to take *te* (although there are differences among the dialects), while other

impersonal forms, such as *musaj* or *si* 'must', do tend to take *te*. Fully inflected modal verbs similarly may or may not be followed by *te*.

Type 1 dialects are regionally diverse, and include dialects from Bulgaria (Yerli, Malokonare, Rešitare), southwest Ukraine, as well as Lithuanian Romani. Here, the complementiser is present with all modals. In these dialects, 'can' is expressed by a borrowed, finite verb (*mogin-/možin-*) or a borrowed impersonal (*može*). A series of dialects (Types 2–5) allow optional use of the complementiser with some modals. In Vălči Dol and Prilep Arli (Type 2), the modal is inherited *našti* which may be followed by *te*. Polish Romani (Type 3) allows both negated *možin-* and *našty*, and borrows 'must', and in Helsinki (also Type 3), 'cannot' is similarly a negated loan verb, whereas 'must' is *mote* (Swedish *måste*) and the absence of the complementiser *te* in this case may be due to avoidance of syllable repetition. Both Sofia Erli and Sípos Rumungro (Type 4) show optional *te* after the inherited *ašti/šaj* 'can' and *našti/nāštik* 'cannot', respectively, whereas in the Gadžikano dialect (Type 5), inherited complement verbs take *te* while Turkish-derived verbs in complement position take the Turkish optative (which figures in Balkan Turkish as a subjunctive in modal complements), with no *te*.

Table 14.2. Complementiser *te* in modal complements with identical subject

	'cannot'	'can'	'must'	'begin'	'try'	'want'
Type 1	*te*	*te*	*te*	*te*	*te*	*te*
Type 2	–/*te*	*te*	*te*	*te*	*te*	*te*
Type 3	–/*te*	*te*	–/*te*	*te*	*te*	*te*
Type 4	–/*te*	–/*te*	*te*	*te*	*te*	*te*
Type 5	–/*te*	–/*te*	–/*te*	–/*te*	–/*te*	*te*
Type 6	–	*te*	*te*	*te*	*te*	*te*
Type 7	–	*te*	–/*te*	*te*	*te*	*te*
Type 8	–	–/*te*	–/*te*	*te*	*te*	*te*
Type 9	–	–	*te*	*te*	*te*	*te*
Type 10	–	–	–	*te*	*te*	*te*
Type 11	–	–	–	–	–	–

The other dialects (Types 6–12) show a hierarchy of absence of the complementiser with certain modals. Here, 'cannot' is usually expressed by inherited *našti/našči/naj* etc. Type 6 includes a large group of dialects primarily from the Balkans (Šumen Drindari, Kaspičan, Crimean, Nange, Kalburdžu, but also Austrian Lovari). In Florina Arli, also of this type, 'cannot' sometimes shows addition of a negator to the complement verb (*našti n-avela khere* 'he cannot come home'). Types 7–8 are sub-types of 6, showing dialects like Muzikanta (Type 7) and Varna Kalajdži (Type 8) where the modals 'can' and 'must' are borrowed, and allow optional insertion of the complementiser. Type 9 is common in the western Balkans and central Europe (Dasikano, Kalderaš, Slovak Romani, Rumungro, Polish Lovari), showing inherited impersonal *šaj* for 'can'. Type 10 comprises the Sinti dialects and neighbouring Roman, but is in effect an extension of Type 9, since here 'must' is the amalgamated form *humte/iste*, which already contains the historical complementiser. Slovene Romani, finally, is the only dialect that employs an infinitive form consistently without a complementiser (Type 11): *sa džanu keri* [all know.1SG do.INF] 'I can do it all').

The types show an interdependency of complexity of the modal complement, and the structure of the modal itself, including person inflection, and borrowings (the latter are more likely to inflect, and to take the complementiser even if they are not inflected). The overall picture is nonetheless one in which action toward a goal ('try', 'begin') and volition ('want') rank higher for complexity than ability ('can' and 'cannot') and necessity ('must').

14.2. Differentiation

This picture fits well with the general tendency of certain modals to be represented by impersonal forms, thus showing less differentiation for person as well as TAM than other modals. The majority of dialects have retained an inherited, impersonal form for 'cannot', and numerous dialects have retained impersonal 'can' as well, as the only forms for this modal. Inherited 'must' is also impersonal, whereas all modals expressing volition, emotion (e.g. 'fear', 'dare'), or action ('try', 'begin') are inflected for person, whether inherited or borrowed. Among the borrowed modals, both 'can' and 'must' may be either person-inflected (as in *možin-*, *biri-* 'can', or *musin-* 'must') or impersonal (as in *može* 'can', or, *lazimi*, *trobuj*, *musaj* 'must'). There are no direct correlations and hence no implications among these individual modals as far as person-inflection is concerned, and many dialects show variation among several

modal forms, usually inherited alongside borrowed forms. Among the 53 core sample dialects evaluated for syntactic structures, 'cannot' is more likely to be impersonal (45 dialects do not allow person inflection, while only 5 require person inflection), as are 'can' and 'must', though by a much smaller margin (for each 30 dialects require impersonal forms, while 20 require personal forms). Thus, there is no ranking between ability and necessity, but both rank lower than volition, emotion or action in respect of differentiation. The inherited modals *šaj/ašti* 'can' and *našti* etc. 'cannot' do not inflect for tense, either, making ability less differentiated for tense than either necessity or volition, emotion, or action.

In some Vlax dialects, active participles of unaccusative verbs (verbs of motion and change of state) as well as mediopassives are used in an evidential meaning (cf. Matras 1995), co-existing with the person-inflected forms, which are neutral in respect of modality. Due to adjectival inflection evidentials are differentiated for gender (M *gelo* F *geli* 'went'), while person-inflected forms are not (*geljas* 'went'). On the other hand, evidentials are restricted to the third person, and so are less differentiated for person than the 'neutral' forms.

14.3. Linear order

Whereas indicative declarative clauses may show either SV or VS order, conditional clauses favour a thetic or predication-oriented presentation that places the event in the foreground:

(1) Karditsa Arli
An dela barbal, na dzaa avri.
if give.3SG wind NEG go.1SG out
'If it is windy, I will not go out.'

(2) Klenovec Rumungro
Te phudla i balval, na džā āri.
if blow.3SG.FUT ART wind NEG go.1SG.FUT out
'If it is windy, I will not go out.'

The order in complements introduced by *te* however is flexible, and subject to the conditions of discreteness. Thus, different-subjects in manipulation clauses may be foregrounded:

(3) Polish Romani
Me kamav kaj jof te džał peske.
I want.1SG COMP he COMP go.3SG REFL.DAT
'I want him to go away.'

(4) Vălči Dol
Mangav vov te džal-tar peske.
want.1SG he COMP go.3SG-away REFL.DAT
'I want him to go away.'

14.4. Borrowing

Expressions of modality rank moderately high among function words that are prone to borrowing (see Table 14.3). Type A includes dialects that have no loans in these functions (e.g. Sinti, Bohemian Romani, and Roman). Type B shows dialects which have borrowed necessity expressions only (e.g. Klenovec Rumungro, West Slovak Romani, Gilan Arli, Sepečides, and Kosovo Bugurdži). Type C (e.g. Welsh Romani, Slovene Romani, Florina Arli, Yerli, Varna Bugurdži, Malokonare, Nange, Muzikanta, Šumen Drindari, Gadžikano) shows borrowings for necessity and ability, while negative ability is encoded by indigenous expressions. In Type D (e.g. Polish and Lithuanian Romani, and Prilep Arli), in addition, negative ability is encoded by negation of the borrowed ability expression. The figure demonstrates a clear implicational hierarchy, with necessity outranking affirmative ability and affirmative ability outranking negative ability: 'must' > 'can' > 'cannot' (see also Chapter 10).

Conditional particles are borrowed into some dialects (Table 14.4). The overwhelming majority of the sample dialects retain inherited *te* (Type A). But borrowings are found in realis conditional clauses (Type B) in Varna Bugurdži,

Table 14.3. Borrowing of modal expressions

	'must'	'can'	'cannot'
Type A	–	–	–
Type B	+	–	–
Type C	+	+	–
Type D	+	+	(+)

Kumanovo Arli (*ako*) and Lithuanian Romani (*jesli*), and for all conditional clauses (Type C) in Velingrad Yerli, Rešitare, Slovene Romani and Muzikanta (*ako*), Karditsa Arli (*an*), Florina Arli and Epiros (*ama*). There thus appears to be a slight tendency for the realis to be favoured for borrowing.

Table 14.4. Borrowing patterns of conditional particles

	Realis	Potential	Irrealis
Type A	inherited		
Type B	borrowed	inherited	
Type C	borrowed		

Chapter 15
Transitivity

Transitivity is clearly definable in terms of the valency of verbs. In Romani, transitivity is recognisable not just by argument structure, but frequently also through valency affixing. Transitivity interacts with the derivational classes of causative, factitive, and middle verbs. *Causatives* are valency-increasing derivations from verbs (e.g. *beš-l-jar-* 'seat' < *beš-* 'sit'); they are transitive by definition. *Factitives* are transitive derivations from adjectives (e.g. *kal-jar-* 'blacken' < *kalo* 'black'). These two derivational classes are marked by one of the Early Romani transitive suffixes **-av-*, **-(j)ar-*, and **-ker-* (see Chapter 5), or by a combination thereof. Causatives and factitives may show identical or distinct marking, depending on dialect. *Middles* are mostly derived by **-jov-*, which alternates with **-il-* in the perfective. Middles derived from verbs function as anticausatives (e.g. *phařav-d-jov-* 'open ITR' < *phařav-* 'open TR') or, more rarely, as passives (e.g. *arakh-l-jov-* 'be found' < *arakh-* 'find').[1] Middles derived from adjectives may be termed *inchoatives* (e.g. *kal-jov-* 'turn black' < *kalo* 'black'). In the Vlax dialects, there are also inchoatives in **-áv-* (e.g. *dil-áv-* 'become crazy' < *dilo* 'crazy'), an intransitive suffix that also survives in adaptation markers of intransitive loan verbs in most dialects. Middle verbs are almost always intransitive,[2] but not all intransitives show middle morphology. We refer to the latter as *active* intransitives. In addition to de-verbal and de-adjectival derivations, there are also transitive and intransitive verbs derived from nouns and adverbs, although these derivations are rather marginal in Romani.

Transitive morphology tends to be more complex and is more likely to extend; it might also be considered more diverse. Intransitives are more exposed through distinct morphology in perfective verbs, and tend to be more differentiated. There is no salient borrowing asymmetry.

15.1. Complexity

Transitive predicates are per definition syntactically more complex than intransitive predicates, as they entail at least two arguments. In most dialects, transitive as well as intransitive verbs only cross-reference the grammatical

subject. However, in a few dialects, the asymmetry in syntactic complexity is also reflected in verb morphology, viz. in marking of pronominal direct objects on transitive verbs. In the Apennine dialects, object inflection of transitives clearly results from fusion of enclitic accusative pronouns with subject-inflected verb forms: e.g. Abruzzian Romani *dikkēmə* < **dikhel ma*(*n*) 's/he sees me', *dikkašt* < **dikhas tu*(*t*) 'we see you.SG'. In Epiros Romani, on the other hand, suffixal marking of third-person direct objects may be a rare retention of a Proto-Romani feature (see Chapter 5): e.g. *dikhav-i* 'I see her', *dikhljom-os* 'I saw him'.

There are several complexity asymmetries in valency-changing morphology. If they are described in terms of the distinction Transitive vs. Intransitive, the asymmetries are in conflict: There are both transitivising and de-transitivising derivations, and so transitives may be more complex than intransitives – if the former are derived from the latter (e.g. the causative *dara-v-* 'frighten' from the intransitive *dara-* 'fear'), or less complex – if the latter are derived from the former (e.g. the middle *šun-d-jov-* 'be audible' from the transitive *šun-* 'hear').[3] Nevertheless, there is some evidence that, despite the overall conflict in the direction of valency-changing derivation, transitives tend to employ more complex valency-changing morphology than intransitives.

In some dialects, the causative derivation may be re-iterated, whereby causatives of causatives (or 'double' causatives) are formed: e.g. Šóka Rumungro *dara-* 'fear' > *dara-v-* 'frighten' > *dara-v-av-* 'make frighten', *an-* 'bring' > *an-av-* 'have brought, order [= make so. bring sth.]' > *an-av-av-* 'have ordered [= make so. make so. bring sth.]'. Double middles, on the other hand, do not occur. This is, of course, partly motivated by the fact that one cannot usually decrease valency of verbs with a single argument. However, middles derived from (active) intransitive verbs do occur in some dialects: they involve aktionsart modification rather than change of valency (e.g. Šóka Rumungro *khand-isaj-ov-* 'stink intensively' < *khand-* 'stink', *ladža-saj-ov-* 'be ashamed constantly' < *ladža-* 'be ashamed'). So it appears that, rather than a ban on middle derivations from intransitives, there is a ban on double application of the middle derivation. In this respect, the intransitive middle morphology is less complex than the transitive causative morphology.

Although in principle the layering of valency-increasing affixes may have a plainly cumulative function in adding new arguments to the predicate structure, as we saw above with 'double' causatives, in practice the modification is often of a lexical-semantic nature. Lexicalisation of valency-increasing morphology has lead to the development of complex markers, which, historically, consist of two transitive markers. The outer marker, and so obviously the

more productive one, is usually **-ker-*.[4] The combination **-(j)ar-ker-* survives in the factitive suffix *-(j)akir-*, which has replaced the simple suffix *-(j)ar-* in the Northeastern dialects (e.g. *kal-jakir-* 'blacken'). Factitive *-(j)aker-* of the same origin is also attested, alongside *-(j)ar-*, in Slovene Romani and some Arli varieties. In several Balkan dialects, the complex marker *-av-ker-* is now the regular causative marker, while simple *-av-* tends to be lexicalised: e.g. Sepečides *phir-* 'travel' > *phir-av-ker-* 'make travel' (vs. *phir-av-* 'lead, carry, drive'). In older German Sinti, we find complex markers with outer *-əv-* (< **-av-*), namely *-ərv-* (< **-ar-av-*) and, more rarely, *-kərv-* (< **-ker-av-*): e.g. *xač-* 'burn ITR' > *xač-əv/ər/ərv-* 'burn TR', *bango* 'crooked' > *banš-kərv-* 'bend TR'. Since 'double' middles do not occur, we do not find any parallel combinations of valency-decreasing markers.[5]

15.2. Differentiation

Intransitives show a greater differentiation in that they are more likely to co-occur with clitic subject pronouns than transitive verbs are. Following the development of a new set of third-person nominative pronouns in *o-*, the original forms in *l-* (which became cliticised) have been retreating in Romani. The majority of dialects have either lost the subject clitics completely or restricted their distribution to the copula and/or non-verbal predications (for further detailes see Chapter 21). Only a few dialects have retained the subject clitics with lexical verbs. In the Sinti dialects and in Roman, the subject clitics may still co-occur with all lexical verbs, irrespective of their transitivity. However, in two dialects of our sample, Klenovec Rumungro and Austrian Lovari, the subject clitics appear to be able to co-occur only with intransitives. Intransitives may occur in two types of constructions, with or without the subject clitics, while transitives only occur in constructions without the subject clitics. Thus, in these dialects, the gradual retreat of the subject clitics has left the intransitives more differentiated than the transitives.

Asymmetries in the formation of adjectival participles are more difficult to evaluate. The ability of a verb to form an adjectival participle contributes to its differentiation in two respects: the verb has an extra form that, in addition, is differentiated for number and gender.[6] Adjectival participles function as *passive* participles (e.g. *mulo ~ muli ~ mule* 'dead [M.SG ~ F.SG ~ PL]') or, when used to express the third-person preterite, as *active* participles (e.g. *mulo* 'he died', *muli* 'she died', *mule* 'they died'). Active participles in the function of the third-person plural preterite are found with all verbs in most dialects. On

the other hand, the use of active participles for the third-person singular preterite is usually confined to intransitive verbs, beginning with verbs of movement and change of state (unaccusatives), and including some other intransitives, and only rarely transitive verbs. Table 15.1 illustrates the distribution of active participles in the third-person singular preterite; "+" indicates obligatory use of active participles, "~" indicates their use alongside person-inflected form, "–" indicates absence of active participles.

Types 1 (Slovene Romani, Varna Bugurdži), 2 (Rumelian Romani) and 3 (Prilep Arli) might be interpreted as archaic distributions, allowing active participles with transitive verbs; or alternatively as analogies, whereby the active participle is copied into transitive verbs. Type 4 is undoubtedly the most prevalent pattern in the Balkans (Arli of Kumanovo, Florina and Karditsa, Epiros, Sepečides, Sofia Erli, Yerli, Crimean Romani, Kosovo Bugurdži, Malokonare, Muzikanta, Nange, Rešitare, Varna Kalajdži), whereby the group of verbs of 'state' may vary across dialects. In Type 5 (Arli of Gilan and Prilep, Priština Gurbet, Ajia Varvara), the pattern is similar, but exclusive use of the participle is confined to middles. The remaining Types 6–9 may be considered transitional varieties, showing the retreat of the active participle outside the core of the Balkan dialects, and toward the central European varieties: Type 6 is the most common (Finnish Romani, Sípos and older Šóka Rumungro, Vălči Dol and Serbian Kalderaš). The participle is somewhat less prominent in Types 7 (Kalburdžu, Taikon Kalderaš) and 8 (Roman, Polish and Austrian Lovari), with minimum appearance in Type 9 (Hungarian Lovari). The third-person singular active participle is lost almost completely in the north-

Table 15.1. Distribution of third-person singular active participles with different types of verbs

	Movement	Middle	State	Transitive
Type 1	+	+	+	~
Type 2	+	~	+	~
Type 3	~	~	~	~
Type 4	+	+	+	–
Type 5	~	+	~	–
Type 6	~	~	~	–
Type 7	+	+	–	–
Type 8	~	~	–	–
Type 9	~	–	–	–
Type 10	–	–	–	–

ern and western dialects (Type 10), cutting right through the South Central group (Klenovec and Nógrád Rumungro, North Central, Sinti, Welsh Romani, Northeastern), though Finnish Romani remains a conservative periphery.

Table 15.1 suggests a hierarchy for the third-person singular active participle formation within intransitive verbs (movement > middle > state), and illustrates most clearly the lower hierarchical position of the transitive verbs in comparison with intransitives. The hierarchy for passive participle formation is almost the opposite. Passive participles are mostly restricted to transitive verbs (e.g. *ker-* 'do, make' > *kerdo* 'done, made') plus a few intransitive verbs of movement and change of state (e.g. *dživ-* 'live' > *dživdo* 'alive', *beš-* 'sit' > *bešto* 'seated, settled'). However, they are usually not found with middle verbs. Although individual dialect types may show greater differentiation of transitive verbs (e.g. Type 10: no third-person singular active participles, passive participles mostly with transitives), the general picture is that of no clear differention asymmetry between transitives and intransitives with respect to the formation of adjectival participles.

15.3. Extension

There is some evidence that transitive markers may extend to intransitive verbs. No developments in the opposite direction are attested. Cross-dialectal evidence suggests that, in Early Romani, adaptation of borrowed transitives involved the Greek suffixes *-Vz-* or *-Vn-* (in the non-perfective) plus the indigenous transitive suffixes *-ker-* and/or *-ar-*, while adaptation of borrowed intransitives involved the Greek aorist suffix *-Vs-* plus the indigenous intransitive suffix *-áv-*; the Greek aorist suffix was also used in the perfective of transitives (cf. Matras 2002: 128–133). Most dialects retain the Early Romani pattern in intransitives. The non-perfective adaptation marker is usually retained as *-Vs-áv-* (e.g. in Crimean Romani and in the Vlax dialects), or changed to *-Vs-jov-* through replacement of the intransitive *-áv-* by the middle suffix *-jov-* (e.g. in the Northeastern and some Balkan dialects). The perfective adaptation marker is almost always retained as *-Vs-a-jl-* (< *-*Vs-áv-il-*, where *-il-* is a perfective suffix), although it sometimes contains a more complex perfective suffix (e.g. *-Vs-a-nd-il-*).

However, in most Central and South Balkan dialects, the transitive Greek-derived *-Vn-* is now used also with intransitives in the non-perfective, and in some Central dialects even in the perfective aspect. While the extension of the transitive suffix *-Vn-* to middle verbs is only indirect, due to their internal

derivation from transitives (e.g. in Šóka Rumungro, the intransitive *ir-in-ď-ov-*, PFV *ir-in-ď-il-* 'turn ITR' is derived from the transitive *ir-in-*, PFV *ir-in-d-* 'turn TR'), its use with active intransitives (e.g. Central *us-in-*, PFV *us-in-d-* 'swim') is probably due to a direct extension.

15.4. Exposition

Intransitive verbs are more exposed than transitive verbs in that there are perfective markers that only attach to intransitive verbs of motion or change of state: *-il-* forms the perfective of middles, as well as of the verbs *av-* 'to come' and *ač-* 'to stay' and a few more, while *-n-*(*d*)-(*il*)- attaches to psych verbs of the type *asa-* 'to laugh' and *dara-* 'to be afraid'. On the other hand, there is no strictly transitive marking in the perfective.

15.5. Internal diversity

The Early Romani inventory of transitive markers was larger than that of intransitive markers. In addition, various dialect-specific developments have increased the cross-dialectal diversity of transitive marking. Thus, western Core Sinti tends to generalise -(*ə*)*v-* (< **-av-*), while eastern Core Sinti tends to generalise -(*ə*)*r-* (< **-ar-*). Complex transitive markers are also dialect-specific to a great extent (see Section 15.1). The two Early Romani intransitive markers, **-jov-* and **-áv-*, have been either both retained, or the latter has been lost, or the middle derivation as such has been replaced by analytic constructions, especially in Core Sinti and Welsh Romani (e.g. Welsh Romani *bārō dža-* 'grow, become big', lit. 'go big' instead of **bar-jov-*). No dialect-specific middle markers have been created. On the whole, transitives appear to be slightly more diverse than intransitives.

15.6. Borrowing

Borrowing of valency-changing markers is rare, restricted to Hungarian loans in Rumungro (South Central) and Hungarian Lovari (North Vlax). In those dialects, where borrowed valency-changing markers are attested, we find both transitive and intransitive loans, so that there appears to be no borrowing asymmetry in transitivity. For example, Šóka Rumungro uses the Hungarian-

derived causative suffix -(*a*)*tat*-, in addition to the indigenous causative suffix -*av*-, with a few pre-Hungarian verbs in /in/ (e.g. *potʲ-atat-in-av-* 'make pay' < *potʲin-* 'pay'); and the Hungarian-derived intransitive suffix -*āz*-, accompanied by the Greek-derived adaptation marker -*in*-, in numerous intransitives derived from nouns (e.g. *trast-āz-in-* 'collect iron' < *trast* 'iron').

Chapter 16
Case and case roles

In this section we discuss asymmetries in three case-related categories: internal case (or Layer I case), external case (or Layer II case) and case roles. The categories of internal and external case constitute inflectional case. The reason why we distinguish internal and external case is the differing structural domains of these categories. While the category of internal case is encoded in all nominals (i.e. in substantivals and adjectivals), the category of external case is only encoded in substantivals (including substantivised adjectivals). This is illustrated in Table 16.1, where the adjective–noun syntagm *terno čhavo* 'young Gypsy lad' is inflected in both numbers. The table and the following discussion reflect the situation in Early Romani and in most dialects.

There are three values of the category of internal case: the nominative, the oblique, and the vocative. The distinction between the nominative and the oblique is encoded (a) in the inflections of the adjective (e.g. *-o* vs. *-e* in the singular)[1] and (b) in the suffixes that immediately follow the inflectional stem of the noun (*-o* vs. *-es-* in the singular, and *-e* vs. *-en-* in the plural). These noun suffixes can be termed *Layer I* case markers (from a diachronic perspective, cf. Matras 1997) or *internal* case markers (from a synchronic perspective). There is an important structural difference between the two nominal word-classes in the morphological status of the oblique: the oblique is a word-form in adjectivals, but a stem (i.e. not a surface construction) in substantivals.[2] The substantival word-forms that are based on the oblique stem, i.e. that contain an

Table 16.1. Internal and external cases

Internal	External	SG		PL	
NOM	–	*tern-o*	*čhav-o*	*tern-e*	*čhav-e*
OBL	ACC	*tern-e*	*čhav-es*	*tern-e*	*čhav-en*
	DAT	*tern-e*	*čhav-es-ke*	*tern-e*	*čhav-en-ge*
	LOC	*tern-e*	*čhav-es-te*	*tern-e*	*čhav-en-de*
	ABL	*tern-e*	*čhav-es-tar*	*tern-e*	*čhav-en-dar*
	SOC	*tern-e*	*čhav-es-sa*	*tern-e*	*čhav-en-sa*
	GEN	*tern-e*	*čhav-es-ker-*	*tern-e*	*čhav-en-ger-*
VOC	–	(*tern-e*)	*čhav-eja*	(*tern-e*)	*čhav-ale*

oblique marker (*-es-* or *-en-*), may be, metonymically, termed oblique case forms. The oblique form of an adjectival is used in agreement with any oblique form of a noun. In other words, adjectivals show deflected case agreement with nouns: they only agree in internal case.

The vocative is encoded by distinct internal case suffixes in nouns (e.g. singular *-eja* and plural *-ale*). There is no vocative with substantival pronouns. The adjectival vocative is marginal (most dialects prefer to employ substantivised adjectival vocatives, e.g. *tern-eja čhav-eja*) and never exhibits a distinct form in those dialects that allow it (e.g. Sípos Rumungro *čor-e rom-eja* 'poor man!' with a form homonymous to the oblique).

The category of external case has six values: the accusative, the dative, the locative, the ablative, the sociative (also called instrumental), and the genitive (also called possessive). The external case suffixes (*-ke* ~ *-ge*, *-te* ~ *-de* etc., see Table 16.1) follow the (internal) oblique suffix. From a diachronic perspective, they have been termed *Layer II* case markers.

The overall number of inflectional cases in nouns is eight (viz. three internal cases, with the oblique being differentiated into six external cases), while there are only two distinct cases in adjectivals. Internal case markers cumulate number and gender, both in substantivals (with the exception of certain pronouns which do not encode number and or gender) and in adjectivals. External case markers are separatist, although each overt external marker has two variants, whose distribution is sensitive to number to a great extent (see Section 16.6 for details).

The third case-related category is the category of *case roles*. We define case roles as grammatical relations and/or thematic roles encoded by inflectional case and adpositions, with the exception of local and temporal case relations.[3] We divide case roles into two sets of values: core case roles and adverbial case roles. *Core* case roles include the values: Subject (canonical transitive or intransitive subject), Experiencer (experiencer or undergoer as non-canonical subject), Predicative (nominal predicate), Object (direct object), Recipient, Possessee (clausal possessee), Possessor (clausal possessor), Adnominal Possessor, and External Possessor (possessor expressed as a core grammatical relation of the verb in a constituent separate from that which contains the possessed item; cf. Payne and Barshi 1999). *Adverbial* case roles include the values: Benefactive, Goal, Comitative, Instrument, Reference ('about'), Source, Material, Origin, Partitive, Reason (including cause and explanation), Comparative (standard of comparison), Equative (standard of equation), Privative ('without'), Substitutive ('instead of'), and Exceptive ('except for').[4] Definitions of some of our case roles combine syntactic and semantic criteria in a

way that partly reflects structures that are general in Romani. The core case roles contain a stronger syntactic component, while the adverbial case roles contain a stronger semantic component.

In the category of internal case, the oblique is the more complex value, while the nominative is more diverse and more likely to extend and to be borrowed. There are conflicting differentiation asymmetries. The vocative shows intermediate complexity. The genitive and the accusative stand out among the values of the category of external case. The genitive is the most complex and the most differentiated value, which is also diverse and likely to erode. The accusative is the least complex value, which is more differentiated and more likely to erode than the other external cases but the genitive. The dative and the sociative are singled out by some criteria, although never as extremes of an asymmetry scale.

In the category of case roles, adverbial case roles are more complex and more prone to renewal and borrowing than core case roles. There is no obvious asymmetry between adverbial and core case roles according to the criteria of extension and extracategorial distribution. Although there clearly are significant asymmetries within core case roles with regard to the criteria of complexity and diversity, and within adverbial case roles with regard to the criteria of complexity, diversity, and borrowing (see below in relevant sections), the dichotomy between core and adverbial case roles turns out to be the major source of generalisations.

16.1. Complexity

In this section, we first discuss complexity asymmetries in the categories of internal case and external case, respectively, and then use them to evaluate complexity asymmetries in the category of case roles.

As for **internal case** marking, the nominative tends to be less complex than the oblique. In nouns, the nominative singular (in all dialects) and the nominative plural (in some dialects) are zero-coded in some inflectional classes, while this is never the case with the oblique stem (e.g. the nominative *rom* 'husband' vs. the oblique stem *rom-es-*). In one inflectional class of adjectivals, viz. the consonantal adjectives, the nominative singular is always zero-coded, while the oblique is overtly marked in some dialects (e.g. the nominative *šukar* 'beautiful' vs. the oblique *šukar-e*). With nouns, the vocative assumes an intermediate position between the nominative and the oblique. Whenever it is distinctly encoded, it is marked by an overt marker (e.g. *rom-eja* 'hus-

band!). However, in some dialects that have retained a distinct vocative with some inflectional classes, the vocative is homonymous to the nominative with other inflectional classes. Thus, overtly marked vocatives may coexist with zero-coded vocatives.

As for **external case** marking, the accusative is the least complex oblique case in that it is zero-coded with regard to the other oblique cases, in Early Romani and most dialects. In other words, it is mostly formed by a zero derivation from the oblique stem (e.g. *romes-* 'husband' > accusative *romes*), while the other oblique cases are formed by suffixing overt external case markers to the oblique stem (e.g. *romes-* > dative *romes-ke*, locative *romes-te* etc.). In some dialects, the accusative singular of masculine substantivals is formed by deletion of the final /*s*/ of the singular oblique stem, and so it is even less complex (e.g. South Central *romes-* > accusative *rome*). The genitive, on the other hand, is the most complex oblique case in that it contains a morphological slot for adjectival agreement with its head noun, i.e. it shows *Suffixaufnahme* (cf. Plank 1995): e.g. *romes-ker-i*, where *-ker-* marks the genitive and *-i* indicates agreement with a feminine noun in the nominative singular. Thus, the genitive inflections are trimorphemic, consisting of an internal oblique marker, the external genitive marker, and an agreement marker (e.g. *-es-ker-i*); the inflections of the dative, locative, ablative and sociative are bimorphemic, consisting of an internal oblique marker and overt external case markers (e.g. dative *-es-ke*); and the accusative inflections are overtly monomorphemic, consisting of an internal oblique marker alone (e.g. *-es*).

The complexity asymmetries in inflectional case are summarised in (1)–(3). The internal case asymmetry is shown in (1) and the external case asymmetry is shown in (2). Both asymmetries are integrated in (3), which is relevant for substantivals:

(1) Oblique > vocative > nominative
(2) Genitive > ablative, sociative, locative, dative > accusative
(3) Genitive > ablative, sociative, locative, dative > accusative > vocative > nominative

We can make use of the complexity asymmetry for inflectional case in substantivals (3) in evaluating complexity asymmetries in **case roles**. We need to add adpositional encoding as clearly more complex than encoding by inflectional case alone. Some semantic relations may but need not be encoded as case roles, depending on a dialect. Instead, they may be expressed in various periphrastic constructions. Case roles whose equivalents are commonly

periphrastic will be considered to be the most complex. The complexity hierarchy of relevant structures is summarised in (4):

(4) a. Periphrastic > adpositional > inflectional
b. Inflectional: genitive > other oblique cases > accusative > nominative

We now turn to discussing complexity asymmetries in individual case roles, starting with core case roles. Subjects, Possessees, inanimate Objects (see Chapter 20 for details), and stative Predicatives were in Early Romani, and still are in the overwhelming majority of dialects, encoded by the nominative. Clausal and External Possessors and animate Objects were encoded by the accusative case, which was also an option for encoding the Recipient and Experiencer case roles. The other option, prevailing in current dialects, was dative encoding. Adnominal possessor was encoded by the genitive case. The complexity hierarchy for Early Romani core case roles is summarised in (5). The alphabetical order reflexts growing complexity of encoding: (a) nominative, the least complex case; (b) split between the nominative and the accusative; (c) accusative, the least complex oblique case; (d) accusative or dative, a more complex oblique case; and (e) genitive, the most complex case.

(5) a. Subject, Possessee, Predicative
b. Object
c. Possessor, External Possessor
d. Recipient, Experiencer
e. Adnominal Possessor

Any of the core case roles has acquired more complex encoding at least in some dialects, at least in particular constructions: those in (5a–b) may be encoded by an oblique case and those in (5b–c) by overt (non-accusative) oblique case markers. Adpositional encoding is found in Adnominal Possessor and Recipient, and less commonly also in Possessor. On the whole, however, later developments have disturbed little of the Early Romani hierarchy. The only salient exception is the reversal of the degree of complexity between Possessor and Possessee in dialects that have developed a specific possession predicate, i.e. the verb 'to have' (see Section 16.6 for details).

Encoding of adverbial case roles in Early Romani may be reconstructed as follows: Substitutive and Exceptive were, in all likelihood, not encoded as case roles. Their semantic equivalents would have been periphrastic constructions, which renders the above roles the most complex. Comparative and Equative were encoded by the particle *sar* 'than, as' (< 'how').[5] Ablative encoding was

an option for Comparative, but not for Equative, especially when the comparee was in the Subject role. Privative was invariantly encoded by the preposition *bi* 'without'. Prepositions were also available for the Reason (*astjal* 'because of') and the Benefactive and Goal (*vaš* 'for') roles. All of these, however, could be encoded by inflectional case as well. Reference, Source, Material, Origin, and Partitive were encoded by the ablative case. These case roles and also Reason and Comparative will be termed the *separative* adverbial case roles. Comitative and Instrument, the *sociative* case roles, were encoded by the sociative case. Goal and Benefactive were encoded by the dative case, and the accusative appears to have been available for Benefactive.

The complexity hierarchy for Early Romani adverbial case roles is shown in (6). The alphabetical order reflects growing complexity: the encoding in (a) was inflectional; in (b–c) inflectional or adpositional (with the Benefactive being encodable by the least complex oblique case); in (d) inflectional or by a particle; in (e) merely adpositional; in (f) only by a particle; and in (g) periphrastic.

(6) a. Comitative, Instrument, Reference, Source, Material, Origin, Partitive
b. Benefactive
c. Goal, Reason
d. Comparative
e. Privative
f. Equative
g. Substitutive, Exceptive

In the current dialects, any adverbial case role may be encoded by an adposition (see Section 16.6). This is the rarest with Comitative and Instrument, and so they are the least complex adverbial case roles.

On a more abstract level, adverbial case roles are clearly more complex than core case roles, in that the former typically rely on the more complex inflectional cases and adpositions, while the latter typically rely on the less complex inflectional cases.

16.2. Erosion

The genitive and the accusative appear to be most prone to erosion among the oblique cases (i.e. in the category of external case). The former may show a radical erosion of its marker: *-ker-* > *-kr-* > *-k-* > *-č-* > 0 (see Section 16.3).

The latter may undergo deletion of the final /*s*/ of the oblique stem, and so it is "less than zero-coded" and marked by a subtractive morphological process in some dialects (see Section 16.1). There appear to be no salient erosion asymmetries in the categories of internal case and case roles.

16.3. Differentiation

The genitive stands out among the external cases in being consistently the most differentiated value. There are conflicting asymmetries between the internal case markers, depending on the cross-cutting category and the word class. The nominative is more differentiated than the oblique: with regard to number in pronouns and adjectivals; with regard to gender in adjectivals; and possibly with regard to inflectional class in all word classes. The oblique, on the other hand, is more differentiated than the nominative: with regard to number in nouns; and with regard to gender in pronouns. There appear to be no salient differentiation asymmetries in the category of case roles, except that Object shows split marking determined by animacy and other categories (see Chapter 20). Table 16.2 summarises the various differentiation asymmetries in inflectional case. We first discuss differentiation of the external cases, and then number, gender, and class differentiation of the internal cases.

There is no differentiation asymmetry of **external** cases with regard to substantival categories: both numbers of a substantival of either gender are encoded with all external cases. However, as the genitive shows Suffixaufnahme, it has its own adjectival (number–case–gender) subparadigm, and is much more differentiated than the other external cases in this respect. For example, any masculine noun has only two dative forms (e.g. singular *romeske*, plural *romenge* 'husband'), while it has up to eight distinct genitive forms (e.g. singular *romesker-o*, *romesker-i*, *romesker-e*, *romesker-a*; plural *romenger-o*, *romenger-i*, *romenger-e*, *romenger-a*).

Table 16.2. Differentiation asymmetries in the category of case

	Number	Gender	Class
External	GEN > other	GEN > other	GEN > ACC > SOC > other
Internal: nouns	OBL > NOM	(NOM <> OBL)	(NOM > OBL)
Internal: pronouns	NOM > OBL	OBL > NOM	NOM > OBL
Internal: adjectivals	NOM > OBL	NOM > OBL	(NOM > OBL)

Table 16.3. Dative and genitive inflections in Kumanovo Gurbet

Noun	DAT	GEN: M.SG.NOM	GEN:F.SG.NOM	GEN:PL.NOM/OBL
M.SG	*-es-**e***	*-es-**0**-o*	*-es-**0**-i*	*-es-**0**-e*
F.SG	*-a-**k'e***	*-a-**k**-o*	*-a-**k'**-i*	*-a-**k'**-e*
PL	*-en-**g'e***	*-en-**g**-o*	*-en-**g'**-i*	*-en-**g'**-e*

Nevertheless, in some dialects of Macedonia and Kosovo, there are genuine categorial splits in external case marking which are not fully explainable on phonological grounds. The splits concern the dative and the genitive. Table 16.3 shows dative and genitive inflections in Kumanovo Gurbet.

The alternation between the singular *-k'e* (dative; < **-ke*)[6] and *-k-/-k'-* (genitive; < **-ker-*) and the plural *-g'e* (< **-ge*) and *-g-/-g'-* (< **-ger-*) has been inherited from Early Romani and is phonologically conditioned. However, the singular suffixes have further split into consonantal variants (*-k'e* and *-k-/-k'-*), which are used with feminine nouns, and variants with a deletion of the consonant (*-e* and zero), which are used with masculine nouns. This gender alternation in the singular is, synchronically, not explainable on phonological grounds. The diachronic scenario is as follows: First, the velars of the case suffixes were palatalised before front vowels, i.e. in the dative (e.g. *-es-k'e*) as well as in the genitive in any agreement form but the masculine singular nominative in *-o* (e.g. *-es-k'-i*, *-es-k'-e* vs. *-es-k-o*). Second, due to consonant cluster simplification, the palatalised velars were lost after the /*s*/ of the masculine singular oblique. This renders the dative **-es-k'e* > *-es-e* and the genitive forms **-es-k'-i* > *-es-0-i* and **-es-k'-e* > *-es-0-e*. The non-palatalised velar of the genitive form *-es-k-o* had been retained at this stage, which is still attested, for example, in the closely related Priština Gurbet. Finally, due to the paradigmatic dominance of the zero marked genitive with singular masculine nouns (in two out of three forms and in five of six agreement contexts), the zero has also been extended to the genitive agreeing with masculine singular nominatives, giving rise to the form *-es-0-o*. The gender split in the marking of the dative and genitive in Kumanovo Gurbet is morphologically conditioned precisely because the extension of the zero has been a morphological rather than a phonological development.

There is a differentiation asymmetry of external cases with regard to inflectional classes, viz. types of substantivals (nouns vs. various pronouns). Disregarding phonologically conditioned alternations, some external cases show distinct markers for different types of substantivals. First, the genitive had four markers in Early Romani: *-ker-/-ger-* in most substantivals, *-inř-* in the

first-person singular pronoun (*m-inř-o* 'my'), *-ir-* in the second-person singular pronoun (*t-ir-o* 'your.SG'), and *-ar-* in the first- and second-person plural pronouns (*am-ar-o* 'our', *tum-ar-o* 'your.PL'). Distinct reflexes of the four markers are retained in the Vlax dialects and the North Central dialects of South Poland and the adjacent areas of Slovakia. All other dialects have reduced the number of genitive markers to three, through extension of that of the first- to the second-person singular pronoun as well (see Chapter 7). Also, numerous dialects have developed clitic or reduced genitive variants of personal pronouns (e.g. *m-ir-o* vs. *m-o* 'my'), thus showing further differentiation of genitive marking. Most dialects retain the two distinct ways of encoding the accusative found in Early Romani: the suffix *-t* in the second-person singular pronoun (*tu-t* 'you.SG') and zero marking in all other substantivals. Some dialects have created a distinction between clitic and full accusative forms of (some or all) personal pronouns; the clitic forms are derived by deletion of the final consonant of the oblique stem (e.g. *man* > *ma* 'me', *amen* > *ame* 'us', *les* > *le* 'him'). The South Central dialects encode the accusative of masculine substantivals by morphological deletion of the final consonant of the oblique stem (e.g. **romes* > *rome* 'husband', **les* > *le* 'him').[7] On the whole, no dialect shows more than three accusative markers. A marginal class differentiation in the sociative is found in the South Central dialects, where most substantivals take *-ha*/*-ca*, with a synchronic deletion of the masculine oblique *-s-* in the singular (e.g. *rome-ha* 'with the husband'), while the person interrogative *ko* 'who' and indefinites derived from it take *-aha*: cf. *kas-aha*, not the regular **ka-ha*. Other oblique cases (dative, locative, and ablative) show a single marker with all types of substantivals.

While all nouns distinguish **number** in the oblique, a few masculine nouns show number neutralisation in the nominative (e.g. *vast* 'hand, hands' but the oblique singular *vastes-* vs. plural *vasten-*) in Early Romani and some dialects (see also Chapter 6). Thus with nouns, the oblique may be more differentiated in number than the nominative. The opposite asymmetry holds for adjectivals. The nominative fully differentiates number in most adjectival classes, while the oblique shows number neutralisation or at least a homonymy between the masculine singular and the plural, depending on dialect. Although third-person pronouns distinguish number in both internal cases, the oblique forms show regular number marking, while number is marked irregularly, and hence more differentiated, in the nominative.

In Early Romani and most dialects, singular third-person pronouns show **gender** differentiation in all cases. However, some dialects have recently lost the gender distinction in the nominative due to convergence with genderless

contact languages (see Chapter 8 for details); gender is still distinguished in the oblique cases (e.g. Vend *ov* 's/he' but *les-* 'him' vs. *la-* 'her'). Adjectivals, on the other hand, tend to show more gender differentiation in the nominative: many dialects have no overt gender distinctions in the oblique of all adjectival classes (cf. the gender-indifferent oblique *bar-e*, but the nominative masculine *bar-o* vs. feminine *bar-i* 'big'). Noun inflection presents a more complex picture. Gender is clearly differentiated in the oblique singular markers (viz. masculine *-s-* vs. feminine *-a-*), but only rarely in the oblique plural, with most dialects showing *-en-* for inflectional classes of both genders. As for the nominative, some markers are gender-specific, while others are shared by inflectional classes of both genders (e.g. a zero inflection in the singular, *-a* or *-ja* in the plural). Thus, for nouns, the hierarchy of gender differentiation (OBL.SG > NOM > OBL.PL) cannot be formulated as an unambiguous asymmetry in case.

As for **class** differentiation in the internal case markers of substantivals, the nominative plural is more differentiated (showing up to ten markers in a dialect) than the oblique plural (with the maximum of four markers and, in some dialects, with a single marker). Both the nominative singular and the oblique singular exhibit significant class differentiation, with roughly the same number of markers. On the whole, the nominative appears to be more differentiated than the oblique in substantivals, although the asymmetry is not very pronounced.

Similarly, adjectivals tend to exhibit greater class differentiation in the nominative than in the oblique. Disregarding the indeclinable consonantal adjectives, there were four major inflectional classes of adjectivals in Early Romani (see Chapter 5): oikoclitic vocalic adjectivals (most adjectives, genitives, and numerous determiners), xenoclitic adjectives, and the deictic classes of demonstratives and the definite article. Adjectival oblique inflections consist of a class marker (viz. xenoclitic *-on-* and deictic *-l-*; there is no class marker in the oikoclitic class) and an oblique suffix cumulating number and gender: the singular masculine and plural *-e* or the singular feminine *-a*. Importantly, the oblique suffixes proper are identical for adjectivals of all inflected classes. Like the oblique forms, the nominative plural also consists of a class marker (deictic *-l-* or *-n-*; there are no class markers in the other classes) and a proper nominative plural suffix. Unlike the oblique forms, the categorial inflection shows class differentiation: *-e* in the oikoclitic class vs. *-a* in the xenoclitic and deictic classes. There are no separatist class markers in the nominative singular, and the difference between adjectival classes is cumulated in the categorial inflections: the masculine has two to three distinct inflections (oikoclitic stressed *-o*, xenoclitic unstressed *-o* and article *o*, and demonstrative *-va*) and

the feminine four inflections (oikoclitic *-i*, xenoclitic *-o*, demonstrative *-ja*, and article *e*). Thus, if one considers only the categorial inflections, the nominative shows greater class differentiation than the oblique. On the other hand, if one takes into account the class markers as well, the class differentiation of the nominative and of the oblique does not show any significant asymmetry. The above description holds for Early Romani and many dialects that retain a similar distribution of adjectival inflections. We will now briefly survey developments that have modified the picture.

One development has increased adjectival class differentiation in the oblique. While many dialects have completely neutralised gender and number distinctions in the oblique through an extension of the singular masculine and plural suffix *-e* to the feminine, originally marked by *-a*, Nógrád Rumungro and Taikon Kalderaš have undergone the neutralisation only in genitives but not in other adjectivals, thus creating secondary class differentiation (cf. feminine oblique *myř-e* 'my' vs. *koř-a* 'blind' in Taikon). In Bunkuleš Kalderaš, genitives have undergone the neutralisation obligatorily (thus showing only *-e* in the feminine), other adjectivals only optionally (thus showing both *-a* and *-e* in the feminine). Hungarian Lovari shows the most complicated distribution of the two inflections in the feminine oblique: *-a* is used with most adjectivals and with genitives of feminine nouns, while *-e* is used only with genitives of masculine nouns (e.g. *dāk-a pheňake* 'to mother's sister' vs. *dadesk-e pheňake* 'to father's sister').[8] Next, two developments have decreased adjectival class differentiation in the nominative: assimilation of the demonstrative inflection and/or of the article inflection to the inflection of oikoclitic adjectivals, through erosion and/or morphological extension.[9] Inflectional assimilation of demonstratives has most frequently occurred in the feminine singular nominative (e.g. *kod-oja* > *kod-i* 'that' like *bar-i* 'big') and less frequently in the masculine singular nominative (e.g. *kod-ova* > *kod-o* like *bar-o*). The nominative plural of demonstratives retains the deictic class marker *-l-* or *-n-*, while the categorial inflection may be assimilated (e.g. *kodo-l-a* > *kodo-l-e* more like *bar-e*).[10] Inflectional assimilation of the article is common in the feminine singular nominative (i.e. **e* > *i*). Assimilation of the nominative plural may be partial, through extension of the oikoclitic categorial inflection (i.e. **ol-a* > *ol-e* > *l-e*), or complete, through further loss of the deictic class marker *-l-* (i.e. *l-e* > *e*).

Finally, some developments have increased adjectival class differentiation in the nominative. First, the oblique suffixes of the oikoclitic vocalic class have frequently extended to the consonantal class of adjectives (e.g. *šukar* > *šukar-e* 'beautiful'). The nominative of consonantal adjectives remains zero-coded,

but the zero now becomes an inflectional variant rather than a mere by-product of the original indeclinability of the class. Second, a few dialects show secondary differentiation in the nominative plural of oikoclitic adjectivals. Thus, Yerli and Malokonare have introduced the suffixes *-a* and *-o*, respectively, with some adjectives but retained the original *-e* with other adjectives (e.g. Yerli *cikor-a* 'small' vs. *bar-e* 'big').

While the increase of class differentiation in the nominative supports the Early Romani asymmetry, the increase of differentiation in the oblique and the decrease of differentiation in the nominative reduce or even remove the asymmetry in individual dialects. Thus, the greater class differentiation of adjectivals in the nominative than in the oblique is only a tendency which may be removed by diachronic developments.

16.4. Extension

In **internal** case, nominative forms or markers extend to the oblique in several instances. In xenoclitic masculine nouns, the vowel of nominative inflections borrowed from Greek may be reconstructed to have extended to the oblique in the Early Romani period: e.g. the original oblique **for-es-* changed to *for-os-* through the influence of the nominative *for-os* 'town, market' (cf. Elšík 2000b). The nominative form of the person interrogative *kon* 'who' has become the base of the oblique stem *kon-es-* in the northern dialects, Slovene Romani and some Drindari varieties, replacing the irregular Early Romani oblique **k-as-*. Demonstrative forms of the masculine singular nominative optionally extend to the oblique (as well as the feminine and/or the plural) in several dialects (e.g. Welsh and Lithuanian Romani, western North Central dialects, or Kalburdžu).

There are also extensions among the **external** cases. In Polish Romani, the ablative extends to the locative, completely replacing its forms, so that original ablative and locative functions are both marked by the ablative forms in *-tyr ~ -dyr*. In the closely related Lithuanian Romani, the merger of ablative and locative functions is underway. It is bidirectional: both sets of functions may be marked by both inflections, the ablative *-tyr ~ -dyr* or the locative *-te ~ -de*. These mergers have been facilitated by functional affinities between the ablative and the locative, as well as by their formal similarity (the case markers share the consonant */t ~ d/*). Several Balkan dialects have changed the sociative case marker **-sa* to *-sar ~ -car* (*-džar*), probably due to influence of the ablative marker *-tar ~ -dar*.

There are numerous instances of extension in case roles, mostly due to convergence with contact languages. Some of them are mentioned in Section 16.6. They do not appear to constitute general asymmetrical patterns.

16.5. Extracategorial distribution

Since we recognise a separate category of *Localisation* for encoding of local and temporal case relations (see Chapter 17), we consider extensions of core or adverbial case markers to local and temporal domains to be instances of extracategorial distribution rather than extension proper. The criterion of extracategorial distribution is relevant primarily to the category of case roles, and only indirectly to inflectional case. Since, however, it is not always discernible which of the case roles that an inflectional case encodes is the source for the extension, we will make use of the names of inflectional cases in quotation marks (e.g. "nominative") to refer indirectly to the relevant case roles. For the range of case roles the individual inflectional cases encode see Sections 16.1 and 16.6. We exclude the locative and the ablative from the current discussion, as they primarily encode local relations.

The "accusative" and the prepositions *astjal* 'because of' and *bi* 'without' do not show any local or temporal uses. The preposition *vaš* 'for' is extended only very rarely. In Slovene Romani, it is attested in the superior localisation (e.g. *vašo kher* 'above the house') and in the temporal simultaneous relation with festivals (e.g. *šu Bušiči* 'on Christmas'). And in some Central dialects, it is attested in the telic extent relation (e.g. Lučivná Slovak Romani *vaš o štar džives* [for ART four day] 'in four days'). Thus, local and temporal extensions are mostly restricted to the "nominative", the "genitive", the "dative", and the "sociative".

The "nominative" in local domains is only found in the Balkans and is restricted to the incorporative localisation with names of towns or countries (e.g. Prilep *ka džas Skopje* 'we will go to Skopje', Kalburdžu *bešel Rusija* 's/he lives in Russia'). In all dialects, the "nominative" is always an option in the atelic extent temporal relation (e.g. Hameln Sinti *oxta divesa* 'for eight days'). The use of the "nominative" in the simultaneous relation is typical of some dialects of the Balkans (e.g. Varna Kalajdži *e rat* 'in the evening, at night', *januari čhon* 'in January [month]', or Kosovo Bugurdži *o nilaj* 'in the summer'), although sporadically it occurs elsewhere, too (e.g. Sinti *mitago* 'at noon'). The "nominative" is unattested with clock time or days of the week, and it is only rarely attested with years.

The "genitive" may be extended only to the temporal simultaneous relation. Numerous dialects of the Balkans (e.g. Florina Arli, Varna Bugurdži, Malokonare, Ajia Varvara, Rešitare, Varna Kalajdži) and Lithuanian Romani use the "genitive" with nouns denoting seasons (e.g. *nilaskoro*, *milasko*, *lynaskiro* 'in the summer'). In dialects of various dialect groups (e.g. Lithuanian Romani, Manuš, Roman, Nange, and Varna Kalajdži), it is also found with (some) days of the week (e.g. *kurkesk*(*er*)*o* 'on Sunday'). Manuš and some Central dialects use it with parts of day (e.g. Manuš *tasarlakro* 'in the morning'), and some Central dialects, Nange, and Lovari for clock time (e.g. Lučivná Slovak Romani *dujengro* 'at two' or Lovari *duj čāsengo* 'at two o'clock').

The local "dative" mostly encodes adessive directive. This extension is attested in Manuš, Roman, Sofia Erli, Kosovo Bugurdži, and Gadžikano. Most examples we have denote human localisations (e.g. Roman *gejom leskere dadeske* [go.PRET.1SG his.OBL father.DAT] 'I went to his father'). Austrian Lovari is attested to use the "dative" in the contact and perlative functions (e.g. *dromenge* 'on roads', *vešenge* 'through forests'). The temporal "dative" is common in the simultaneous relation. Numerous dialects of the Balkans employ it with nouns denoting festivals (e.g. *Ramazanoske* 'during Ramadan'), and some of them also with parts of day (e.g. Nange *evlijake* 'in the morning'). The Sinti dialects have specialised the dative for clock time, although Austrian Sinti has extended it to festivals and seasons as well (e.g. *oxtengi* 'at eight', *Nevo Beršeski* 'on New Year's Eve', *herbsteski* 'in the autumn'). An extensive use of the "dative" in the simultaneous relation is also found in Polish Romani (e.g. *belvelake* 'in the evening', *sobotake* 'on Saturday', *vendžake* 'in the winter'). The "dative" may also encode atelic extent: in Welsh Romani and some Arli-type dialects and Erli (e.g. *jekhe diveseske* 'for a day'); or telic extent and future distance: in Roman, Prilep Arli and Kosovo Bugurdži (e.g. *jekhe diveseske* 'in a day').

The local "sociative" is used in the perlative localisation in numerous dialects (e.g. Welsh Romani, Polish and Russian Romani, the Central dialects, or Taikon Kalderaš), due to convergence with Slavic languages (e.g. Taikon *le vəšesa* 'through the forest'). Some Central dialects also extend the "sociative" to the sequentive localisation (e.g. Slovak Romani *le paňeha* 'along the river'). As for temporal extensions, the "sociative" is commonly used in the simultaneous relation in the dialects of the Balkans, especially with nouns denoting parts of day (e.g. *ratjasa* 'at night, in the evening'),[11] and in some dialects also with nouns denoting seasons (e.g. Ajia Varvara *milajesa* 'in the summer'). Another temporal extension of the "sociative" in some dialects of the Balkans (e.g. Florina Arli, Sofia Erli, Varna Bugurdži, Malokonare, and Varna Kalajdži)

Table 16.4. Extensions of core and adverbial case roles into the local and temporal domains

	NOM	GEN	DAT	SOC	*vaš* 'for'
Inessive	*yes*	no	no	no	no
Directive	no	no	*yes*	no	no
Contact	no	no	*yes* (rare)	no	no
Superior	no	no	no	no	*yes* (rare)
Perlative	no	no	*yes* (rare)	*yes* (freq.)	no
Sequentive	no	no	no	*yes* (rare)	no
Simultaneous	*yes*	*yes* (freq.)	*yes* (freq.)	*yes*	*yes* (rare)
Future distance	no	no	*yes*	*yes* (rare)	no
Telic extent	no	no	*yes*	*yes* (rare)	*yes* (rare)
Atelic extent	*yes* (freq.)	no	*yes*	*yes*	no
Case roles	3	1	7	6	3

is the atelic extent in the expression *sahatenca* (etc.) 'for hours'. Telic extent and future distance encoded by the "sociative" are attested in Slovak Romani (e.g. *duje ďivesenca* 'in two days').

The various local and temporal extensions of the core and the adverbial case markers are summarised in Table 16.4.

The "dative" and the "sociative" are the most versatile cases in terms of extension to different localisations and temporal relations. The "nominative" shows medium versatility, as does the preposition *vaš* 'for', where however all extensions are cross-dialectally rare. The "genitive" undergoes a single but frequent type of extension. The asymmetry of core and adverbial case roles with regard to local and temporal extensions may be formulated as follows: "dative" (especially Recipient) > "sociative" (Comitative and Instrument) > "nominative" > "genitive" (Adnominal Possessor), Benefactive (*vaš*) > "accusative", Reason, Privative.

16.6. Internal diversity

In the category of **internal** case, the nominative is more diverse and more likely to be renewed than the oblique. This asymmetry can be exemplified with the diversity of nominative plural markers in nouns (see Chapter 6); or with the renewal of nominative forms in personal pronouns (see Chapter 5) and demonstratives (see Chapter 8).

External case suffixes of most substantivals show mostly only phonological shape variation. In Early Romani, the suffixes beginning in an obstruent developed a morphophonological alternation: after a nasal (viz. after the oblique plural suffix *-en-* and after the irregular oblique stem *man-* of the first-person singular pronoun) the initial obstruent of the suffixes was voiced, while it remained voiceless in all other environments.[12] For simplicity's sake, we call the voiceless forms *singular* variants, and the voiced forms *plural* variants. The sociative suffix *-sa*, beginning in a fricative, was uniform in Early Romani. Dialectal reflexes of the Early Romani external case suffixes are shown in Table 16.5.

In most dialects, the alternations in external case suffixes are phonologically conditioned, whether the difference between the singular and plural variants is voicing of the initial consonant of the suffix or more.[13] Nevertheless, as discussed in Section 16.3, some dialects of Macedonia and Kosovo have developed gender splits in genitive and dative suffixes, which contribute to a greater internal diversity of the genitive and the dative (viz. some dialects show only phonologically conditioned alternations here, while other dialects also show morphologically conditioned alternations).

Internal cross-dialectal diversity is much greater in the adverbial **case roles** than in the core case roles. Innovations in both sets of case roles are frequently due to convergence with contact languages: for example, the extension of the dative to External Possessor; the extension of the dative to Source in numerous dialects in contact with Slavic; the extension of the sociative to Predicative and of the locative to Possessor in Russian Romani; and many more. At least in some instances, convergence is also responsible for the introduction of local adpositions into various adverbial case roles (e.g. *katar ~ tar* 'from' or *andar* 'out of' into most of the separative case roles, and various local adpositions such as *upre* 'on' or *pal* 'behind' into the Reason, Goal, Benefactive, and Reference case roles). Core case roles, on the other hand, are only

Table 16.5. External case suffixes: dialect forms

	SG	Reflexes	PL	Reflexes
DAT	*-ke*	*-ki*, *-kī*, *-kə*, *-t'e*, *-če; -e*	*-ge*	*-gi*, *-gī*, *-gə*, *-d'e*, *-dže*
LOC	*-te*	*-ti*, *-tī*, *-tə*	*-de*	*-di*, *-dī*, *-də; -ne*
ABL	*-tar*	*-tra*, *-tyr*, *-tər*, *-ta*, *-tē*	*-dar*	*-dra*, *-dyr*, *-dər*, *-da*, *-dē; -na*
SOC	*-sa*	*-sar*, *-ha*, *-(j)a*, *-ra*	*-sa*	*-sar*, *-ca(r)*, *-ča(r)*, *-dža(r)*
GEN	*-ker-*	*-kir-*, *-kər-*, *-kar-*, *-kVr-*, *-kr-*, *-k-*, *-t'ir-*, *-č-; -Vr-*	*-ger-*	*-gir-*, *-gər-*, *-gar-*, *-gVr-*, *-gr-*, *-g-*, *-d'ir-*, *-dž-*

rarely encoded by local adpositions (e.g. Adnominal Possessor by *katar ~ tar* 'from' and Recipient by *ke* 'to'). Among the adverbial case roles, Reason, Goal and Benefactive as well as Substitutive and Exceptive appear to be the most diverse, while Comparative and Equative, the separative case roles, and especially Comitative, Instrument and Privative are relatively stable. Among the core case roles, Possessor, External Possessor and Recipient appear to be more prone to renewal than Subject, Experiencer, Object, Predicative, and Possessee.

An interesting instance of cross-dialectal diversity in the encoding of Possessor and Possessee stems from diversity in constructions of clausal possession. In Early Romani and in most dialects, the accusative Possessor and the nominative Possessee are connected by the copula:[14] e.g. [*minře dades*]=ACC *si* [*duj phenja*]=NOM 'my father has two sisters'. In a few dialects (e.g. Sepečides, Prilep and Karditsa Arli, and Abruzzian Romani), there is an alternative construction with reverse case marking. Here, a specific verb of possession *ther-* 'have' (< 'hold') has developed, which behaves as a transitive and assigns the nominative to Possessor and the accusative to an animate Possessee: e.g. [*miro dad*]=NOM *therel* [*duje phenjen*]ACC 'my father has two sisters'. The transitive possessive construction also exists in dialects that borrow the verb 'to have' (e.g. Lithuanian Romani *majin-*).

16.7. Borrowing

As for **internal** case, nominative markers are frequently borrowed, while oblique markers never are (or only rarely).[15] Nominative markers of both numbers were borrowed at the Early Romani stage from Greek as part of Greek noun and adjective loans (e.g. nouns: singular *-os*, *-is*, *-i*, *-a* and plural *-i*, *-a*, *-es*; adjectives: the singular *-o* and the plural *-a*). The Greek singular nominative markers later became adaptation markers that integrated post-Greek loans into xenoclitic inflectional classes (see Chapter 23 and Elšík 2000b), while the Greek nominative plural markers in nouns were, in some dialects, replaced or supplemented by borrowings from later contact languages (e.g. *-uri* from Rumanian, *-ovi* from Bulgarian). Further, case-insensitive plural markers from Turkish and Hungarian were selectively borrowed into the nominative, but not into the oblique forms, of third-person pronouns in a few dialects (e.g. Gadžikano *on-lar* or Nógrád Rumungro *ón-k* 'they'; see also Chapter 6).

None of the **external** case suffixes have been replaced by borrowed case affixes. Nevertheless, nouns may be borrowed in their source case forms. The

case-inflected loans that are borrowed always function as (mostly local or temporal, see also Chapter 17) adverbials in the source language, and they are adverbs rather than noun forms in Romani. For example, Kaspičan and Gadžikano borrow the Turkish locative and ablative forms in *-dA* and *-dAn* respectively (e.g. *Rusija-da* 'in Russia', *Sofija-dan* 'from Sofia'); some dialects of the Balkans borrow the Turkish instrumental forms in *-ile* (e.g. Rešitare *saba-jle* 'in the morning'); and Rumungro borrows from Hungarian the inessive forms in *-bA*[16] (e.g. *idegen-be* 'in a foreign country', *mājuš-ba* 'in May'), the superessive forms in *-Vn* (e.g. *kedd-en* 'on Tuesday'), the temporal forms in *-kor* (e.g. *tavas-kor* 'in the spring'), and the ablative forms in *-tū* (e.g. Romungro *kiškor-tū* 'since the childhood'). The source case markers may show limited productivity as adverbial markers in Romani, being used with some indigenous adverbial expressions (e.g. Rumungro *khēral-tū* 'from home', *idžal-tū* 'since yesterday'), but they never extend to all nouns to become fully productive, inflectional, case markers. Thus, borrowing of source case markers is considered to be borrowing of adverbial markers. In other words, apart from the nominative (see above), there is no borrowing of inflectional case markers in Romani.

As for **case roles**, only borrowing of adpositional case markers into adverbial case roles is attested. There are no loans into core case roles. In addition, there is no borrowing of *distinct* case markers of Source, Material, or Origin. However, these case roles as well as Partitive, Reference, and Reason can be encoded by borrowed local adpositions, since they are, at least in part, metaphorically based on various local case roles (localisations, see Chapter 17). If the relevant localisation shows borrowing of a case marker, the loan can also be used in the above non-local case roles (e.g. in Austrian Sinti, the separative adessive preposition *fon* 'from' of German origin is also used in the metaphorically separative roles of Source, Origin, Partitive, Reference, Reason, and probably also Material).

Borrowing of (non-local) Partitive, (non-local) Reference, Benefactive, and Goal adpositions is rare. A loan Partitive preposition is only attested in Welsh Romani: *o* from English (or possibly Welsh, cf. Sampson 1926: IV, 249). Sofia Erli rarely uses the Bulgarian preposition *za* in the Reference function. In modern Sinti dialects in current contact with German, *fir* is used in the Benefactive and Goal functions and *um* in the Goal function. Manuš has borrowed *fir* (*fur*) from German, an old L2, and *pur* from French, the current L2; both prepositions are only attested in the Benefactive. Abruzzian Romani has borrowed the Benefactive and Goal preposition *pri* from Italian. Gadžikano and Kaspičan make use of the Turkish postpositions *ütürü* and *ičin*. The former

covers the Reference, Benefactive, and Reason case roles, while the latter encodes Goal and sometimes also Reason.

Borrowed (non-local) adpositions of Reason are common in the Balkans, but rare elsewhere. They include: the South Slavic preposition *radi* (*zaradi*, *poradi*) in Slovene Romani, Prilep and Kumanovo Arli, Yerli, and Rešitare; the Bulgarian preposition *za* in Varna Bugurdži; the Macedonian preposition *zbog* in Prilep Arli; the Turkish postposition *ütürü* in Gadžikano and Kaspičan; and the Turkish postposition *ičin* in Gadžikano and Vălči Dol.[17] Only borrowing of the preposition *vegn* from German takes place outside of the Balkans, viz. in German and Austrian Sinti.

Borrowed sociative (Comitative and Instrument) prepositions are well attested. The source languages include Greek (*me*), South Slavic (*s*, *so*), German (*mit*), and Italian (*kun*, *ku*, *ki*). The source language is mostly the dialect's current L2, with the exception of French Manuš and Venetian Sinti, which retain a loan from German, an old L2. Only Italian (Piedmontese, Lombardese, Venetian) Sinti and Apennine (Abruzzian, Calabrian) Romani use the loan prepositions exclusively; this is due to the fact that inflectional cases, including the sociative, have been lost in these dialects, at least with nouns. Other dialects use the loan prepositions alongside the indigenous sociative: the loans are common in Core Sinti, but relatively rare in Arli of Gilan, Kumanovo, Florina and Karditsa, and Ajia Varvara. In Varna Bugurdži and Yerli, the use of the Bulgarian preposition is triggered by Bulgarian determiners (e.g. *s nekolko gostenca* 'with some guests'). In Ajia Varvara and Karditsa Arli, on the other hand, the Greek preposition triggers the use of the Greek definite article if the noun phrase is definite (e.g. *me ta star čhaja* 'with the four girls'). In many dialects, the borrowed prepositions encode both sociative case roles, Comitative and Instrument. Only the Comitative function is attested in Florina Arli, Yerli, Varna Bugurdži, and Ajia Varvara. This seems to indicate that Comitative is more prone to borrowing than Instrument. Also, a specifically Comitative preposition *sos*(*v*)*e* 'together with' (from Macedonian) is borrowed into Kosovo Bugurdži. The Rumanian preposition *ku* 'with' in North Vlax is only attested in Comitative idioms (especially in *ku-sa* 'with all').

Borrowed Privative adpositions are well attested. The source languages include Greek (*xoris*), Slavic (*bez*, *brez*), German (*ōne*, *oni*), and Swedish (*utān*). The Slavic forms frequently trigger contamination of the indigenous *bi* 'without' (e.g. *biz*, *bri*). The source language is mostly the current L2, except for Finnish Romani, where the preposition from Swedish (a recent L2) can be shifted to postpositional position due to convergence with Finnish (the current L2). Only a few dialects completely replace the indigenous preposition:

Finnish and Slovene Romani, and Arli of Prizren, Gilan, Kumanovo and Karditsa. More commonly the indigenous preposition is only supplemented by the loan: in some Sinti varieties and Roman, some varieties of Slovak Romani (e.g. Lučivná, Humenné region), Epiros, Sofia Erli, Yerli, Varna Bugurdži, Xoraxane, and Kalburdžu.

Borrowing of adpositions is the norm with the Substitutive and Exceptive case roles. Many dialects only borrow adpositions of these and no other non-local and non-temporal functions (e.g. Polish and Lithuanian Romani, Muzikanta, Nange, and Malokonare). Slavic is the major source of Substitutive prepositions (e.g. *mesto*, *misto*, *vmesto*, *namesto*, *zamjast*, *mjesta* 'instead of'). They are found in many Northeastern, many Central, most Balkan, and many Vlax dialects, and in Slovene Romani, mostly originating in the dialects' current L2's. Rumungro retains the form *misto* from Croatian, an old L2.[18] Karditsa Arli borrows the Greek Substitutive preposition *andi*, and Kaspičan uses the Turkish-derived postposition *jerine*. Borrowed Exceptive prepositions, including complex prepositions, are numerous. They include: *ektos* or *ektos apo* from Greek in Epiros, Karditsa Arli and Ajia Varvara; *osven* or *s isključenie* from Bulgarian in numerous dialects of Bulgaria; *pokraj* from Macedonian in Kumanovo Arli; *osim* (*sem*) from Serbian in Kosovo Bugurdži, Serbian Kalderaš, and Dasikano; *okrem* and *krom'e* from Slovak and Russian, respectively, in Slovak and Lithuanian Romani; and *opruč* or *z vyn'ontkem* from Polish in the Northeastern and Central dialects of Poland.[19]

There are only few borrowing implications in the category of case roles. A loan of a Goal or a Benefactive adposition appears to imply a loan of a Reason adposition, which in turn implies the encoding of the Substitutive and Exceptive case roles through a borrowed adposition or periphrasis (but not through an indigenous adposition). Borrowing of Reference is independent of Reason loans. There is no implicational asymmetry between Reason loans on the one hand, and loans of Privative or sociative adpositions on the other hand. Some dialects possess only the former (e.g. Kaspičan, Gadžikano, Rešitare, or Vălči Dol), while others possess only (one or both of) the latter (e.g. Helsinki Romani, some Slovak Romani varieties, Roman, Arli of Florina and Karditsa, or Kalburdžu).[20] There is no implicational asymmetry between borrowing of Privative and sociative adpositions either. Some dialects possess only a Privative loan (e.g. Helsinki Romani, Roman, Slovene Romani, Epiros, or Kalburdžu), while other dialects possess only a sociative loan (e.g. Italian Sinti, Apennine Romani, Florina Arli, or Ajia Varvara).[21] Although borrowed Substitutive and Exceptive adpositions are much more common than borrowed Privative adpositions, there is no strict implicational asymmetry:

cf. Helsinki Finnish Romani, which borrows a Privative adposition but makes use of an indigenous preposition for the Substitutive and Exceptive case roles. There are three common patterns in Substitutive and Exceptive adpositions: either none is borrowed; or both are borrowed; or the Substitutive adposition is a loan, while periphrasis is used to encode the Exceptive role. Only Sofia Erli appears to use periphrasis with the Substitutive role, but has a loan in the Exceptive role. Although the Partitive is one of the least likely adverbial case roles to be borrowed, it may be the only loan of a non-local and non-temporal adposition in a given dialect.

Chapter 17
Localisation

The category of localisation is encoded in local adverbials: local noun phrases, local adpositional phrases, and local adverbs.[1] It is a multivalue category. Different values of the category, different localisations, express distinct spatial positions of a figure object with regard to a ground object.[2] In local phrases, localisation is encoded through case marking: by inflectional case markers in local noun phrases (e.g. *veš-este* 'in [a] forest') and/or by adpositions in local adpositional phrases (e.g. *andr-o veš* 'in the forest'). The referent of the head noun of a noun phrase or the referent of the object noun of an adpositional phrase is the ground object of localisation in local phrases. In local adverbs, localisation is encoded by the lexeme itself (e.g. *andre* 'inside') or more precisely by its stem (see below). Localisation encoded in adverbs does not allow an explicit, overtly coded, ground object. Instead, the ground object of adverbial localisation is implicit, being retrievable from context (situation, discourse, or shared knowledge).

The values of the category of localisation that may be distinctly encoded in Romani, together with broad paraphrases of their semantics, are shown in Table 17.1 (reading: the *inessive* localisation expresses the position of a figure object *inside of* a ground object, and so on).

Localisation values may be divided into four groups. *Core* localisations express the basic spatial configurations: containment of the figure object within a closed space of the ground object (inessive), an explicit lack of containment (extraessive), adjacency between the two objects (adessive), and contact of the figure object with the surface, especially the top surface, of the ground object (contact). The proximate, which indicates that the figure object pertains to the immediate sphere of the ground object (without focusing on, but not excluding, their adjacency or contact), often being located beside or next to the ground object, and the distant, which encodes spatial distance, are subsumed under *proximity* localisations.[3] *Axis* localisations consist of the superior and inferior localisations in the vertical dimension, and of the anterior and posterior localisations in a horizontal dimension. Finally, there are several *peripheral* localisations. Table 17.1 does not contain the *prolative*, a case function with non-local as well as peripheral local semantics found in constructions such as *pull by hair*.

Table 17.1. Localisation values in Romani

Localisation			Meaning
Core		INESSIVE	*inside of*
		EXTRAESSIVE	*outside of*
		ADESSIVE	*at*
		CONTACT	*on the surface of*, *on the top of*
Proximity		PROXIMATE	*by*, *in the surroundings of*
		DISTANT	*far from*
Axis	*vertical*	SUPERIOR	*over*, *above*
		INFERIOR	*under*, *below*
	horizontal	ANTERIOR	*in the front of*
		POSTERIOR	*behind*, *in the back of*
Peripheral		MEDIAL	*between*, *among; in the middle of*
		OPPOSITIVE	*opposite*, *on the opposite side of*
		TRANSLATIVE	*across*, *over*
		PERLATIVE	*through*
		CIRCUMLATIVE	*around*
		SEQUENTIVE	*along*, *past*

Importantly, distinctions in orientation (e.g. stative, directive, separative) do not constitute different localisation values. Instead, orientation is considered to be a separate category (see Chapter 18). In some localisations, there are two or more expressions per each relevant structure (inflectional case, adposition, adverb) which encode the same spatial configuration but differ in their orientation, i.e. in whether the configuration is encoded as an actual one (stative orientation), one to be assumed through movement of the figure object (directive orientation), or an original one abandoned through movement (separative orientation). Orientation is thus, conceptually, a cross-cutting category in respect to localisation. This is also confirmed by separate marking of either category. For example, the stative/directive inessive adverb *andr-e* 'inside, inward' and the separative inessive adverb *andr-al* 'from inside' share the stem *andr-* (which marks the inessive localisation) and differ in the suffixes *-e* vs. *-al* (which are the markers of the stative/directive vs. separative orientations, respectively).[4] There are some important differences between case markers and adverbs with regard to which localisations they may encode. Localisations marked by inflectional case are discussed in detail in Section 17.1. With the exception of the extraessive and the distant localisations, which are rarely encoded as case relations (and even if they are, they show some structural

Table 17.2. Early Romani local adpositions

Localisation	Orientation	Form	Translation
Inessive	Stative/directive	*andre*	'in, into'
	Separative	*andar*	'out of'
Adessive	Stative/directive	*ke ~ te*	'at, to'
	Separative	*katar ~ tar*	'from'
Contact–superior	Stative/directive	*opre*	'on; over'
(Contact)	Separative	*opral*	'from the top of'
Proximate		*paš*	'by'
Inferior		*tel*	'under'
Anterior		*angle*	'in front of'
Posterior		*pal*	'behind'
Medial		*maškar*	'between, among'
Oppositive		*mamuj*	'opposite'
Translative–perlative		*perdal*	'across, over; through'
Circumlative		*trujal*	'around'

anomalies and clearly result from dialect-specific innovations), all localisations may be encoded by adpositions. Table 17.2 charts local adpositions as they are reconstructed for Early Romani. Note that separative adpositions can only be reconstructed with some certainty in the core localisations. Also note that the (stative/directive) superior localisation was probably homonymous to the contact localisation, and that there was no distinction between the translative and the perlative. It is not clear which adposition marked the sequentive localisation, and so it is not included in the figure.

Local adverbs do not encode the adessive and the contact localisations. Sequentive, translative, perlative and circumlative adverbs are rarely attested in our data (if they are, they mostly correspond to adpositions in form), and so we do not include them in our discussion or in Table 17.3, which shows the local adverbs reconstructed for Early Romani.[5]

The core localisations are the least complex, the most likely to erode, the most differentiated, and perhaps the most likely to extend; they are also relatively diverse but not very likely to be borrowed. The axis localisations are the least diverse and the least likely to be borrowed; they are relatively complex and do not show much differentiation. The peripheral localisations, on the other hand, are the most diverse and the most likely to be borrowed, the least

Table 17.3. Early Romani local adverbs

	Stative/directive		Separative	
Inessive	*andre*	'inside, inward'	*andral*	'from inside'
Extraessive	*avri*	'outside, outward'	*avrjal*	'from outside'
Proximate	*paše*	'nearby'	*pašal*	'from nearby'
Distant	*dur*	'faraway'	*dural*	'from faraway'
Superior	*opre*	'above'	*opral*	'from above'
Inferior	*tele*	'below'	*telal*	'from below'
Anterior	*angle*	'in/to the front'	*anglal*	'from the front'
Posterior	*pale*	'in/to the back'	*palal*	'from the back'
Medial	*maškare*	'in the middle'	*maškaral*	'from the middle'
Oppositive	*mamuj*	'on/to the other side'	*mamujal*	'from the other side'

differentiated and are relatively complex. The proximate localisation does not assume any salient position: it is relatively complex, not very differentiated, not very diverse, and quite borrowable.

The adessive is the most prominent value within core localisation: it is the most differentiated, the most likely to extend, the most diverse, and the most likely to be borrowed. The inessive is the least likely to extend and the least diverse; it is likely to erode and shows an intermediate degree of differentiation. The contact localisation is the least differentiated value, which is slightly more complex than the other core values; it is likely to erode and shows an intermediate tendency towards extension and intermediate diversity.

As for the axis localisations, the horizontal localisations are more complex and more likely to extend than the vertical localisations. The vertical localisations, on the other hand, are more likely to be borrowed. There is conflicting evidence with regard to differentiation: while the horizontal localisations are more differentiated in local case markers, the vertical localisations are more differentiated in local adverbs. Within the horizontal localisations, the posterior is more likely to extend, while the anterior is more differentiated and more diverse. Within the vertical localisations, the superior is more diverse, while the inferior is more differentiated.

Within peripheral localisation, the oppositive is the most complex value, which is also most likely to be borrowed. The medial is the least diverse value, which is also least likely to be borrowed. The perlative appears to be the least complex peripheral localisation. The translative, the circumlative, and the sequentive are not assigned any prominence by our criteria.

17.1. Complexity

In order to be able to evaluate complexity asymmetries in the category of localisation with regard to **case marking**, we have to assume a complexity hierarchy of the structures (case markers) that encode localisation. This structural complexity hierarchy is given in (1).

(1) a. adpositional > inflectional
b. adpositional (type of adposition): complex > simple
c. adpositional (case governed): oblique cases (other > accusative) > nominative
d. inflectional: oblique cases > nominative

Adpositional encoding is clearly more complex than inflectional encoding (a), as it involves a free case marker, and circumpositions or prepositional groups consising of two adpositions (complex adpositions) are clearly more complex than simple adpositions (b). Since oblique cases (accusative being the least complex oblique) are more complex than the nominative (see Chapter 16), adpositional phrases governing obliques and inflectional obliques are more complex case markers than adpositional phrases with a nominative and adpositionless nominatives, respectively (c–d).

The localisations that may be encoded by inflectional case include: inessive, adessive, contact, perlative, prolative, and marginally also medial, translative, and circumlative. There are no local uses of the accusative or the genitive. The least marked inflectional case, the nominative, is restricted to the non-separative inessive ('in, into') with names of towns or countries in a few dialects of the Balkans (e.g. Prilep Arli, Gadžikano, Varna Kalajdži, Kalburdžu); see the example in (2):

(2) Kalburdžu
Mere duj pheja bešen ***Švecija.***
my two sisters sit.3PL Sweden(.NOM)
'My two sisters live in Sweden.'

The locative is very frequent in the non-separative inessive ('in, into'), less so in the non-separative adessive ('at, to'). Both functions are attested in Welsh Romani, the South Central dialects, numerous South Balkan dialects (e.g. Arli of Prilep and Florina, Sepečides, Erli, Yerli, Varna Bugurdži), Kosovo Bugurdži, and Ajia Varvara; see the examples in (3). Only the inessive

locative is attested in some Balkan dialects (e.g. Malokonare, Muzikanta, Nange, Gadžikano) and in Rešitare. Only the adessive locative, on the other hand, is found in Slovene Romani, and some Vlax dialects (e.g. Lovari, Taikon Kalderaš, Varna Kalajdži). The adessive locative is mostly employed with animate objects; see the example in (4).

(3) Sepečides (Cech and Heinschink 1999: 107)
a. *Kerasa buti ekhe gaves-**te.***
work.1PL work one.OBL village-LOC
'We work in a village.'
b. *Ekhe masa-**te** bešle.*
one.OBL table-LOC sat.3PL
'They sat down at a table.'

(4) Taikon Kalderaš (Gjerdman and Ljungberg 1963: 106)
*Gəlo-tar peskə dades-**te.***
went.3SG-away REFL.GEN father-LOC
'He went to his father's place.'

The locative may also be used in the non-separative contact localisation 'on' (e.g. in Prilep Arli, Kosovo Bugurdži, Austrian Lovari, or Ajia Varvara); see the example in (5). A perlative ('through') use of the locative is only attested in Sofia Erli (6).

(5) Ajia Varvara (Igla 1996: 104)
*Pošuja-**te** pašliasas.*
floor-LOC lie.1PL.REM
'We used to sleep on the floor.'

(6) Sofia Erli
*Našti tradav kalke dromes-**te.***
cannot drive.1SG this.OBL road-LOC
'I cannot drive through this road.'

The distribution of the ablative partly parallels that of the locative. It is found in those core localisations as the locative, except that it is specialised for the separative orientation:[6] i.e. inessive 'out of', adessive 'from', and contact 'from the top/surface of'.[7] All three functions of the ablative are attested in

the Northeastern dialects, Varna Bugurdži, and Nange; see the examples in (7). In some dialects, one of the three core localisations cannot be encoded by the ablative.[8] Instead, adpositional encoding must be used: in the inessive in Gadžikano (*an*); in the adessive in Kalburdžu (*katar*); and in the contact localisation in Šóka Rumungro (*pal* or *upral*) and Slovene Romani (*zuro*). In some other dialects, the adessive is the only core localisation that can be encoded by the ablative (e.g. in Sofia Erli and Varna Kalajdži).

(7) Polish Romani
   a. *Jov vygeja štarybnas-**tyr.***
      he out.went.3SG jail-ABL
      'He got out of jail.'
   b. *Łeskry phen' javja targos-**tyr.***
      his sister came.3SG marker-ABL
      'His sister came from the market.'
   c. *Da flaśka speja tyśa-**tyr.***
      that bottle off.fell.3SG table-ABL
      'That bottle fell down from the table.'

The ablative is also employed in a number of peripheral localisations. The translative ('across, over'), circumlative ('around'), and especially perlative ('through') functions are well attested in the Balkans, viz. in some Balkan dialects (e.g. Yerli and Varna Bugurdži) and in the South Vlax Kalburdžu. The prolative function ('by') of the ablative is more widespread, being also attested in some Sinti, Central, and North Vlax dialects. We exemplify ablative encoding of peripheral localisations from the Nange dialect, which is the only dialect of our sample that also has a medial ('between') ablative (8):

(8) Nange
   a. *Me nakom o mostes-**tar.***
      I passed.1SG the bridge-ABL
      'I went across the bridge.' (translative)
   b. *Oj phiradas pes amare keres-**tar.***
      she walked.3SG REFL.ACC our.OBL house-ABL
      'She walked around our house.' (circumlative)
   c. *Naši nakava kale dromes-**tar.***
      cannot pass.1SG this.OBL road-ABL
      'I cannot drive through this road.' (perlative)

d. *Ov sidijas eke čavra o balen-**dar.***
he pulled.3SG one.OBL girl.ACC ART hair.ABL
'He pulled a girl by her hair.' (prolative)

e. *Nanaj but baro rastojanie o masa-**tar** dži karik mo leglos.*
is.not much big space ART table-ABL up.to towards my bed
'There is not much room between the table and my bed.' (lit. 'from the table up to my bed' (medial)

The dative encodes the directive adessive ('to, toward'), mostly with animate objects, in several dialects (e.g. Manuš, Roman, Kosovo Bugurdži, Gadžikano); see the example in (9). Austrian Lovari shows a perlative ('through') use of the dative (10).

(9) Roman (Halwachs 1998: 89)
*Idž leskere dades-**ke** gejom.*
yesterday his.OBL father-DAT went.1SG
'Yesterday I went to his father's place.'

(10) Austrian Lovari (Cech and Heinschink 1998: 38)
*Žav bare dromen-**ge**, vešen-**ge.***
go.1SG big.OBL roads-DAT forests-DAT
'I go through big roads, through forests.'

Finally, the sociative is, due to convergence with Slavic, commonly used to encode the perlative ('through') localisation (e.g. in Welsh Romani, the Northeastern and the Central dialects, and Taikon Kalderaš). Sociative encoding of the sequentive ('along') is restricted to a few Central dialects. Examples in (11) are from Klenovec Rumungro:

(11) Klenovec Rumungro

a. *Čhāve dikhnahi la hevja-**ha** and-o ploto.*
children see.3PL.REM the.OBL hole-SOC in-the fence
'The children looked through the hole in the fence.'

b. *Phirasahi upre tēle le pāňi-**ha.***
walk.1PL.REM up down the.OBL water-SOC
'We were walking up and down along the river.'

There is mostly an alternative adpositional encoding of the localisations that may be encoded by inflectional case. The factors that determine the choice between inflectional and adpositional encoding include animacy, definiteness, determination, referentiality, lexicality (pronouns vs. nouns), proprriality (proper vs. common nouns) and other sorts of prominence, or their more or less complex interplay, depending on dialect. The only localisations that may exhibit exclusive inflectional encoding are the core localisations (inessive, adessive, contact) in the separative orientation. There is no adpositional alternative for any of these localisation, in Lithuanian Romani, for example, Varna Bugurdži, and Nange, and further dialects show exclusive inflectional encoding for at least some of the core localisations. Exclusive adpositional encoding, i.e. no inflectional alternative, is found with the proximate, vertical (superior and inferior), horizontal (anterior and posterior), and oppositive localisations.

In adpositional encoding of a localisation, there is mostly a split between adpositional constructions with the object noun in an oblique case (mostly locative, but also the accusative or the ablative) and adpositional constructions with the object noun in the nominative. For the sake of convenience, we will use the terms *oblique vs. nominative* (*adpositional*) *constructions*. This split is determined by a selection from the same set of factors as the split between inflectional and adpositional encoding (see above); an overwhelming majority of dialects retain oblique constructions at least with personal pronouns. Provided there is adpositional encoding at all, different localisations behave alike, i.e. they allow both oblique and nominative constructions under identical conditions. The only exception is the adessive preposition *ke* ‘at, to’, which requires the nominative even with personal pronouns in the Northeastern dialects (with the exception of Polish Romani) and in Welsh Romani. Contrast the nominative construction in the adessive localisation (12a) with the oblique construction in the proximate localisation (12b) in Estonian Romani:

(12) Estonian Romani
   a. *Jekh džukel jawdža* ***ke jöj.***
      one dog came.3SG to she(.NOM)
      ‘A dog came to her.’
   b. *Jekh džukel jawja* ***paš late.***
      one dog came.3SG by her.LOC
      ‘A dog came (close) to her.’

The South Balkan dialects of Prilep, Florina and Karditsa Arli, Sepečides, Yerli, Varna Bugurdži, and Crimean Romani, and the Balkan *zis*-dialect Nange have developed a system of complex adpositions. In all of these dialects, the complex adpositions consist of an adposition that is specific to the given localisation and the adessive adposition *ke ~ te* 'at, to' (the former in most of the dialects, the latter in Florina Arli, both in Nange). The latter thus functions as a general local case marker. The complex adpositions are mostly preposed to their object noun or, especially in Sepečides, they are occasionally circumpositions, with the localisation-specific adposition postposed due to Turkish influence (e.g. *talal k-o rukh* or *k-o rukh talal* 'under the tree'). The complex adpositions alternate with simple adpositions: the former generally occur with the nominative, while the latter occur with dependent oblique cases; see the examples in (13).

(13) Crimean Romani
   a. *Pujos garavel pes* ***pal*** ***ko*** *škaf'i.*
      child hide.3SG REFL.ACC behind at cupboard(.NOM)
      'The child is hiding behind the cupboard.'
   b. *Ov sos' garavelas* ***pal*** *pe dumeste.*
      he something hide.3SG.REM behind REFL.GEN back.LOC
      'He was hiding something behind his back.'

The complex adpositions mostly occur in all localisations, either only in the non-separative orientation (e.g. Varna Bugurdži *an ke* 'in, into' but only *andar*, not **andar ke* 'out of'), or in all orientations (e.g. Crimean Romani *ande ke* 'in, into' and *andar ke* 'out of'). The complex adpositions may also contain borrowed localisation-specific adpositions (e.g. Yerli *is ke* 'out of, from', *mežđu ke* 'between', or Florina Arli *karšia te* 'opposite'). The major exception is the adessive localisation. The stative/directive adessive is always encoded by a simple *ke ~ te*; the adposition is never doubled (e.g. **ke ke*). However, directive and separative adessive may be encoded by a complex adposition (e.g. Crimean Romani *kar'in ke* 'towards', *katar ke* 'from'). The inessive, like the other localisations, allows complex adpositions, with two exceptions: In Florina Arli, where the adessive prepositon *te* also encodes the inessive (e.g. *t-i Florina* 'in Florina'). And in Nange, where the simple inessive preposition *ande* 'in, into' alternates with the adessive preposition *te* specified by the adverb *andre* 'inside'.

The complexity asymmetry among localisations with regard to case marking is summarised in (14):

(14) a. adessive, inessive;
b. contact;
c. perlative;
d. translative, circumlative, sequentive, medial;
e. proximate, vertical, horizontal, oppositive

The adessive and the inessive (a) are the least complex localisations, since they are most frequently encoded by inflectional case. The adessive is less complex than the inessive in that it may require the nominative in adpositional constructions (which is never the case with the inessive and other localisations); and in that it does not allow complex adpositions in the stative/directive orientation (which is only rarely the case with the inessive). The inessive, on the other hand, may be encoded by the least complex inflectional case, the nominative. The contact localisation (b), together with the adessive and the inessive, need not allow adpositional encoding, which is never the case with the other localisations. Inflectional encoding is relatively common with the localisations in (c), and possible with those in (d). Finally, the localisations in (e) are always encoded by adpositions, and hence most complex.

Peripheral localisations are more complex in that they are usually not encoded in local **adverbs**: for example, there is rarely a circumlative adverb 'around'; instead an adpositional phrase with a dummy object would have to be used. Among the core and axis localisations, complexity asymmetries may arise through the so-called ablative shift, whereby ablative forms in *-al* extend from the separative orientation to the stative and/or directive orientations (see Chapter 18). The ablative shift is most likely to affect horizontal adverbs followed by vertical adverbs, and it is least likely to occur with core adverbs. For example in Slovak Romani of Lučivná, the shift affects only the horizontal localisation (*angl-al* 'in/to the front', *pal-al* 'in/to the back') but not the vertical or core localisations (*upre* 'above', *tele* 'below', *andre* 'inside', *avri* 'outside'). And in older German Sinti, the shift affects, in the stative orientation, the vertical as well as horizontal localisations (*pr-al* 'above', *tel-al* 'below' and *gl-an* 'in the front', *pal-al* 'in the back'), but not the core localisation (*drin* 'inside', *vrin* 'outside').[9] Horizontal adverbs are thus more likely to be more complex than vertical adverbs, which are in turn more likely to be more complex than core adverbs.

17.2. Erosion

Erosion affects both local adpositions and local adverbs, more the former than the latter (e.g. *pe* 'on' vs. *opre* 'above' in numerous dialects). With the exception of the highly erodable inessive and contact prepositions (cf. *andre* 'in, into' > *are*, *ane*, *no*, *dro*, *do*; *opre* 'on' > *pre*, *pe*, *ap* and *opral* 'from the surface of' > *pra*, *pa*), there does not seem to be any clear erosion asymmetry among different localisations. In local adverbs, monosyllabic forms may develop in various localisations: e.g. core *dran* (< **andral*), vertical *pral* (< **opral*), horizontal *glan* (< **anglal*) in Sinti.

17.3. Differentiation

The major differentiation criterion, relevant to both case marking and adverbs, is the number of distinctions in the cross-cutting category of orientation (see Chapter 18). Localisations differ with regard to encoding of the separative orientation as a **case** relation. First, the separative is generally encoded in the inessive ('out of'), adessive ('from the point of'), and contact ('from the surface of') localisations by specific adpositions or the inflectional ablative. Second, case encoding of the proximate, inferior, and horizontal separatives is attested only in a few dialects. In Šóka Rumungro, they are encoded by specific ablative adpositions: proximate *paš-al* 'from the position near, next to' (vs. *paš* 'near, next to'), inferior *tal-al* 'from the position under' (vs. *tal* 'under'), anterior *angl-al* 'from the front of' (vs. *angle* 'in/to front of'), and posterior *pal-al* 'from the back of' (vs. *pal* 'behind'). In Ajia Varvara, the anterior separative is encoded by a non-separative adposition with ablative dependent case marking, while in the stative/directive the same adposition governs the locative (e.g. *angla man-dar* 'from the front of me' vs. *angla man-de* 'in front of me'). Similarly in Florina Arli, anterior orientation is encoded by differential dependent case marking with a single complex adposition (e.g. *angla te man-dar* 'from the front of me' vs. *angla te man-de* 'in front of me').[10] Third, superior, oppositive, and medial separatives are never encoded as case relations. Finally, the separative orientation is incompatible with the other peripheral relations.

Non-separative case markers (adpositions and inflectional cases) generally do not encode the distinction between the stative and the directive orientations. There are two exceptions, however. First, in the adessive localisation, numerous dialects possess a specific directive and/or a limitative directive case

Table 17.4. Orientation distinctions in case marking by localisation

Localisation	Separative	Directive	Orientations
Adessive	always	frequently	2–3
Inessive	always	very rarely	2–3
Contact	always	never	2
Anterior	rarely	never	1–2
Posterior, inferior, proximate	very rarely	never	1–2
Superior, peripheral	never	never	1

marker (i.e. allative 'to, towards' or 'up to'): the dative in various dialects (e.g. Roman, Kosovo Bugurdži, or Gadžikano), the preposition *karig* (*karing* etc.) in many Balkan and Vlax dialects as well as in Latvian Romani, or loan adpositions in Sinti (the German-derived prepositions *gegn*, *nax*, *bis*) or Rumungro and Lovari (the Hungarian-derived postposition *felē*). Second, in the inessive localisation, some Central dialects of Slovakia possess a specific directive (i.e. illative) case with names of localities (e.g. *Požom-u* 'to Bratislava').

Table 17.4 summarises the differentiation asymmetries in case localisations with regard to the frequency and the number of orientation distinctions. The overall hierarchy is: adessive > inessive > contact > anterior > posterior, inferior, proximate > superior, peripheral.

Core **adverbs** tend to be the most differentiated in terms of number of orientation distinctions they encode, followed by vertical adverbs; horizontal adverbs show the least differentiation. This is illustrated from Šóka Rumungro (Table 17.5). While the core adverbs distinguish four orientations – viz. the directive, the stative, and two separatives: a static one (=1) and a dynamic one

Table 17.5. Orientation distinctions in core and axis local adverbs in Šóka Rumungro

		Directive	Stative	Separat.1	Separat.2
Inessive	'inside'	*ānde*	*onďānde*	*āndral*	*āndraltū*
Extraessive	'outside'	*āri*	*onďāri*	*āvral*	*āvraltū*
Superior	'up'	*upre*	*upral*	*upraltū*	
Inferior	'down'	*tēle*	*tēlal*	*tēlaltū*	
Anterior	'in the front'	*ānglal*		*ānglaltū*	
Posterior	'in the back'	*pālal*		*pālaltū*	

(=2) (see Chapter 18 for details) – the vertical adverbs have a single separative, and the horizontal adverbs, in addition, do not distinguish the directive and the stative orientations.

17.4. Extension

In this section, we only discuss extensions affecting case relations encoded by indigenous adpositions. In example glosses, we give the original meanings of extending adpositions. There does not seem to be any clear extension hierarchy among the four localisation groups, except perhaps that the core localisations extend most commonly. There is, however, a clear hierarchy among the core localisations (adessive > contact > inessive) and a clear hierarchy among the axis localisations (posterior > anterior > vertical). Vertical localisations are the only localisations that never extend.

Core localisations never extend to non-core localisations, with the exception of the generalisation of the adessive adposition as a general local marker in complex adpositions (see Section 17.1). Among themselves, extensions of any core value to any other core value are attested. However, there are important asymmetries in frequency and type of different extensions. **Adessive** extensions (i.e. extensions of adessive adpositions) are the most frequent, especially in the dialects of the Balkans. The Kumanovo Arli examples in (15) illustrate the extension of the adessive adpositions *ke* 'at, to' and *kotar* 'from' to inessive (ab) and contact (cd) localisations.

(15) Kumanovo Arli
   a. *Me živinavaine* ***k****-o* *baro kher.*
      I live.1SG.REM at-ART big house
      'I used to live in a big house.'
   b. *Oj iklili* ***kotar*** *o* *kher.*
      she came.out.3SG.F from ART house
      'She came out of the house.'
   c. *O lil tano* ***k****-o* *astali.*
      ART leaf be.3.M at-ART table
      'The letter is on the table.'
   d. *O šiši pelo* ***kotar*** *o* *astali.*
      ART bottle fell.3SG.M from ART table
      'The bottle fell down from the table.'

Different types of adessive extension may be distinguished according to which orientations are affected, and according to whether the original inessive or contact adpositions have been completely replaced or not. Table 17.6 shows the different types of extensions of adessive adpositions to the inessive localisation. The pluses and minuses, respectively, indicate the presence or absence of individual adpositions (viz. extension or non-extension of adessive adpositions, and retention or loss of inessive adpositions) in the inessive localisation. Partial extension is marked by lighter shading, complete extension by darker shading.

Type A, with no adessive-to-inessive extension whatsoever, is found in most dialects outside of the Balkans as well as in a few Balkan and South Vlax dialects (e.g. Prizren Arli, Priština Gurbet, Vălči Dol, and Kalburdžu). This type represents the Early Romani state.

In Types B-D, the non-separative adessive adposition *ke ~ te* 'at, to' extends to the inessive localisation (completely replacing the original inessive adposition *andre* 'in, into' in Type D), but there is no adessive extension in the separative orientation. The original separative inessive adposition *andar* 'out of' is lost Types C and D, through generalisation of the inflectional ablative or

Table 17.6. Adessive and inessive adpositions in the inessive localisation

	Non-separative		Separative	
	andre	*ke ~ te*	*andar*	*katar ~ tar*
Type A	+	–	±	–
Type B	+	+	+	–
Type C	+	+	–	–
Type D	–	+	–	–
Type E	+	–	+	+
Type F	+	–	–	+
Type G	+	+	+	+
Type H	+	+	–	+
Type I	–	+	–	+

through replacement by a loan adposition, i.e. it is not lost due to adessive extension. Type B is found in some varieties of Slovak Romani, Crimean Romani, Muzikanta and Rakarengo, where the adessive adposition is rare and appears to be restricted to certain object nouns (e.g. to names of localities in Slovak Romani: ***kij**-e Moskva* 'in Moscow'), as well as in Kosovo Bugurdži. In this latter dialect, there is a specialisation of the adessive adposition for directive orientation, while the original inessive adposition encodes the stative orientation (e.g. ***k**-o kher* 'into the house' vs. ***an**-o kher* 'in the house'). Extension of the adessive adposition is rare and lexically restricted also in Type C (Slovene Romani, Varna Bugurdži and Nange). On the other hand, in Type D (Yerli), the adessive *ke* completely takes over the inessive localisation.

In Types E and F, there is an adessive extension in the separative, but not in the non-separative orientation. The separative adessive adposition *katar ~ tar* 'from' is a rare alternative to the original inessive adposition *andar* 'out of' in Type E (Lovari, Taikon Kalderaš and Ajia Varvara). In Type F (Montana Kalajdži), on the other hand, the inessive adposition is lost through replacement by the adessive adposition.

In Types G-I, adessive extension affects the inessive localisation of both orientations. The inessive adpositions are still more frequent in Type G (Malokonare, Rešitare and Varna Kalajdži). Sofia Erli is somewhat special in that in the non-separative orientation, the inessive adposition is more frequent than the adessive one, while in the separative orientation the opposite asymmetry holds. In dialects of Type H, the separative inessive *andar* 'out of' is lost, while the non-separative inessive *andre* 'in, into' is retained. In some dialects of this type (Gadžikano and Kaspičan), the separative inessive adposition is mostly lost through decomposition of localisation and orientation marking (see Section 17.6), but there are also rare extensions of the adessive adposition *katar* 'from'. In other dialects of this type (viz. Prilep Arli and Kumanovo Gurbet), the separative inessive is lost through replacement by the adessive *katar ~ tar* 'from', and the non-separative inessive is retained as a very rare alternative to the adessive *ke* 'at, to'. Finally, in Kumanovo Arli (Type I), the adessive adpositions *ke* 'at, to' and *kotar* 'from' have completely replaced the original inessive adpositions.[11]Extensions of adessive adpositions to the contact localisation are likewise common. In some dialects (e.g. Kumanovo Arli and Gurbet, Montana Kalajdži, and Malokonare), the adessive adpositions *ke ~ te* 'at, to' and *katar ~ tar* 'from' have completely replaced the original contact adpositions *opre* 'on' and *opral* 'from the surface of'. In other dialects (e.g. Varna Kalajdži, Rešitare, and Sofia Erli), there is still variation in the non-separative orientation. In Velingrad Yerli and Nange, there is a complete

extension of the non-separative adessive adposition, while separative contact is encoded by the inflectional ablative or by a loan adposition.

Extensions of the non-separative **contact** adposition *opre* 'on' to the inessive and adessive localisations are attested in numerous dialects. However, they are mostly restricted to constructions with certain object nouns, for example to localities in case of contact-to-inessive extensions (e.g. Finnish Romani *ape themenne* 'in villages', Estonian Romani *po fōros* 'in the town', Sípos Rumungro *upre jekhe fōroste* 'in a town'). Typically, convergence with contact languages is involved in the choice of adposition (e.g. Polish Romani uses the contact adposition in *pre gav* 'in the village' but the inessive adposition in *ando foro* 'in the town', calquing Polish *na wsi* and *w mieście*, respectively). Extension of the separative contact adposition to other core localisations is only attested in Kalderaš and Gadžikano. Examples (16a) and (16b) illustrate, respectively, the inessive use and the contact use of the preposition *pa* (< **opral*) 'from the surface of' in Serbian Kalderaš:

(16) Serbian Kalderaš (Boretzky 1994: 114–115)
a. *Dikhel* ***pa*** *e feljastra.*
see.3SG from.surface ART window
'S/he looks out of the window.'
b. *Avile* ***pa****-v* *istoko maj godźaver manuš.*
came.3PL from.surface-ART east more wise men
'The wisest men came from the east.'

Extensions of **inessive** adpositions are the rarest among the core localisations. Examples include the extension of the separative adposition *andar* 'out of' to the adessive localisation with a few object nouns (e.g. Muzikanta *andar o pazari* alongside *katar o pazari* 'from the market'), and the extension of the non-separative adposition *andre* 'in, into' to the contact localisation; see the example in (17):

(17) Slovene Romani
Čhide leske ***do*** *gra.*
put.PRET.3PL him.DAT in horse
'They put him on a horse.'

A summary of extensions among core localisations is given in Table 17.7. Extending adpositions are identified in the left half of the figure, localisations that are extended upon are given in the right half.

Table 17.7. Extensions among core localisations

Localisation	Orientation	Adposition	Adessive	Contact	Inessive
Adessive	non-separative	*ke ~ te*		common	common
	separative	*katar ~ tar*		common	common
Contact	non-separative	*opre*	common		common
	separative	*opral*	rare		rare
Inessive	non-separative	*andre*	–	rare	
	separative	*andar*	rare	–	

Although all extensions within the core localisations are attested, adessive extensions are cross-dialectally common, while contact and especially inessive extensions are much rarer. Moreover, some inessive extensions are not attested in certain orientations (viz. inessive-to-adessive extensions in the non-separative and inessive-to-contact extension in the separative). Finally, extending adessive adpositions may completely replace adpositions of the contact and the inessive localisations, while the latter never extend completely. The extension hierarchy among the core localisations is adessive > contact > inessive.

There are numerous dialects where certain examples containing the **proximate** adposition *paš* (*pašal*) 'by' may be interpreted as encoding the adessive localisation. However, since the definition of the proximate localisation (the figure object pertains to the immediate sphere of the ground object) does not exclude the definitional feature of the adessive localisation (the two objects are adjacent), it is difficult to interpret such examples as instances of a proximate-to-adessive extension. Rather, proximate encoding reflects an alternative conceptualisation of spatial configurations that may also be encoded as adessive: while adessive encoding explicitly expresses adjacency, proximate encoding leaves it out of focus (cf. the Estonian Romani examples in 12, which are alternative responses to an identical questionnaire sentence). It may be no accident that most examples that are interpretable as adessives and encoded by the proximate adposition involve human ground objects, where adjacency is mostly metaphorical anyway, and hence where there is less need for it to be explicitly encoded; see the example in (18).

(18) Varna Kalajdži
Lesko dad bičalel les ***paša*** *leski dej.*
his father send.3SG him.ACC by his mother
'His father sends him to his mother's place.'

Nevertheless, the proximate-to-adessive extension has, no doubt, taken place in Sinti. While most dialects retain an overall distinction between the proximate *paš* 'by' and the adessive *ke ~ te* 'to, at', modern Sinti dialects have completely replaced the latter by the former; see the example in (19). As mentioned above, there is no extension in the opposite direction (i.e. adessive-to-proximate), and so we may formulate the extension hierarchy proximate > adessive.

(19) Hungarian Sinti (Mészáros 1980: 19)
džin ***paš*** *u gap*
LIM by ART village
'up to the village'

The proximate adposition *paš* (*pašal*) 'by' may also extend to the superior localisation (e.g. in the Northeastern dialects and Kalburdžu) and/or to the anterior localisation (e.g. in Lithuanian Romani and Nange); see the examples in (20). Some of the extensions appear to be systematic (e.g. there is no alternative to *paš* in the superior localisation in the Northeastern dialecs).

(20) Lithuanian Romani
a. *Partr'eto visinel* ***paš*** *e ložko.*
portrait hang.3SG by ART bed
'A portrait hangs above the bed.'
b. *Me užakira tut* ***paš*** *kxangiri.*
I wait.1SG.FUT you.ACC by church
'I will wait for you in front of the church.'

There is no extension of vertical localisations. Horizontal adpositions may extend to the inessive and the contact localisations, especially to encode the separative orientation of these localisations. Extensions of the **posterior** adposition *pal* 'behind' are common in the Central dialects. The Slovak Romani examples in (21) illustrate the posterior extension to the separative inessive (a), the non-separative inessive (b), the separative contact (c), and the non-separative contact (d). The extension to separative contact localisation

is also found in the South Central dialects, where *pal* alternates with the original *up*(*r*)*al* (< **opral*) 'from the surface of', perhaps partly due to their formal similarity. On the other hand, the extensions of *pal* to the non-separative orientation of the two localisations are restricted to a few Central varieties and to certain ground objects or constructions.[12]

(21) Slovak Romani (Lučivná)
  a. *Vičinel* ***pal*** *e blaka.*
  call.3SG behind ART window
  'S/he calls out of the window.'
  b. *Somas* ***pal*** *but štati.*
  be.PRET.1SG behind many states
  'I have been in many countries.'
  c. *Avľom tele* ***pal*** *o graj.*
  came.1SG down behind ART horse
  'I came down from the horse.'
  d. *Dikhľom le muršes*, *sar džal* ***pal*** *o drom.*
  saw.1SG ART man.ACC how go.3sg behind ART road
  'I saw the man walking on the road.'

The **anterior** adposition *glan* (< **anglal*) 'in/to the front of' is only attested as extending to the separative inessive ('out of') in Austrian Sinti; see the example in (22).

(22) Austrian Sinti
Har van o Sinti ***glan*** *o lagera.*
how came.3PL ART Sinti in.the.front.of ART camp
'wHen the Sinti came out of the camp.'

Finally, there are numerous extensions of peripheral adpositions. There does not seem to be any clear hierarchy among them. The medial preposition *maškar* 'between' extends to the perlative function in a variety of dialects (e.g. Estonian Romani, some Sinti, Šóka Rumungro, Muzikanta, and Nange), and to the sequentive meaning 'past' in Nange. The oppositive preposition *mamuj* 'opposite' extends to the translative function in a few dialects of Bulgaria (e.g. Muzikanta, Varna Kalajdži, and Rešitare). The translative-perlative preposition *perdal* 'across, over; through' extends to the sequentive meaning 'past' (but not 'along') in Lučivná Slovak Romani, and to the circumlative and superior functions in Šóka Rumungro. And the circumlative preposition *tru-*

jal 'around' extends to the sequentive meaning 'along' in a few dialects of Bulgaria (e.g. Sofia Erli, Velingrad Yerli, Varna Bugurdži, and Kalajdži).

17.5. Extracategorial distribution

Localisation markers are not only used to encode spatial configurations. They are also used metaphorically, to encode temporal and other abstract relations between objects. We consider the metaphorical use of localisation values to represent their extended, extracategorial distribution. In this section, we discuss the distribution asymmetries among different localisations with respect to temporal metaphors. Our classification and terminology of temporal relations is based on Haspelmath (1997a).

Several localisation hierarchies resulting from an application of the criterion of distribution to temporal metaphors are shown in (23): (a) a hierarchy among the four localisation groups; (b) a hierarchy among peripheral localisations; (c) a hierarchy among non-separative core localisations; (d) a hierarchy among separative core localisations; and (e) a hierarchy among axis localisations. Certain salient non-temporal metaphors (the separative case roles, see Chapter 16) are mostly based on the separative core localisations, and so they confirm the widest extracategorial distribution of the core localisations.

(23) a. Core > axis > proximate, peripheral
b. Peripheral: medial > other
c. Core (non-separative): inessive > adessive > contact
d. Core (separative): adessive > other
e. Axis: horizontal > vertical

There are four extremely common temporal metaphors (which the dialects appear to have inherited together from Early Romani), and a few less common ones (which appear to be the result of dialect-specific internal innovations or calquing). The major temporal metaphors are shown in (24):

(24) a. Inessive (or other core) → simultaneous
b. Adessive (limitative) ('up to') → anterior-durative ('until')
c. Adessive separative ('from') → posterior-durative ('since')
d. Horizontal ('behind') → sequence ('after') and distance ('in X's time')
e. Horizontal ('in front of') → sequence ('before') and distance ('ago')

As the first major metaphor, non-separative core case markers are used to encode simultaneous temporal relations. Among these, **inessive** markers are the norm, while the use of adessive and contact markers in the temporal domain reflects their extensions to the inessive localisation in the local domain (see Section 17.4).

In Early Romani, simultaneous relation with parts of days, days of weeks, seasons, and possibly months was encoded by the so-called old (Layer I) locative in *-e ~ -i* (e.g. *javin-e* 'in the morning', *dives-e* 'during the day', *belvel-e* 'in the evening', *rat-i* 'at night'; *kurk-e* 'on Sunday'; *nilaj-e* 'in the summer' and *jevend-e* 'in the winter'). The old locative has been retained in numerous dialects, and even extended to borrowed nouns (e.g. Slovak Romani *sombat-on-e* 'on Saturday', *jar-on-e* 'in the spring', Polish Romani *styčn'-on-e* 'in January'). Although the old locative is, synchronically, a de-substantival adverb derivation rather than an inflectional case, we may consider it to be an inessive marker. The current (Layer II) inflectional locative in *-te ~ -de* is rarer in simultaneous relations, and mostly restricted to parts of days and clock time (e.g. Welsh Romani *diveses-tī* 'during the day', Sepečides *akšamis-te* 'in the evening', Varna Kalajdži *efta saxaten-de* 'at seven o'clock').

The inessive adposition *andre* 'in, into', on the other hand, can be used in any type of simultaneous relation, most commonly with months and years (e.g. Polish Romani *ando čerfco* 'in July', *ando vavir berš* 'in the last year'). In Welsh and Polish Romani, and in numerous dialects of the Balkans, the preposition has been extended to seasons (e.g. Kosovo Bugurdži *ano nilaj* 'in the summer'), and in Slovene Romani and some Northeastern dialects to days of the week and clock time (e.g. Slovene Romani *nu sreda* 'on Wednesday', *nu štar* 'at four'). Many dialects use it with various parts of days (e.g. Yerli *andivin* < **andi javin* 'in the morning', Lithuanian Romani *do pašdyves* 'at noon', Sípos Rumungro *andi rāt* 'at night').

The adessive adposition *ke ~ te* 'at, to' encodes simultaneous temporal relations first of all in those dialects where it has replaced the inessive adposition in local uses (e.g. Arli of Kumanovo, Prilep and Florina, and Yerli): cf. Prilep Arli *ko efta o saati* 'at seven o'clock', *ko kurko* 'on Sunday', *ko vend* 'in the winter', *ko duj hiljade i biš* 'in [the year] 2020', *ko Ramazan* 'during Ramadan'. However, the adessive may be also used in dialects where the extension of the adposition *ke* to the inessive is rare in local uses (e.g. Sofia Erli, Varna Bugurdži, Kosovo Bugurdži, Nange, Rešitare); then it is usually restricted to a certain type of simultaneous relation. Extensions of the contact adposition *opre* 'on' to the inessive localisation likewise licence its occasional use for simultaneous relations: most examples are with parts of days (e.g. Slovak

Romani *pro dilos* and Kalburdžu *po mesmeri* 'at noon', Austrian Sinti *ap i rat* and Slovene Romani *po rače* 'at night'), and some with seasons and festivals (e.g. Kalderaš *pe primovara* 'in the spring', Klenovec Rumungro *pi karāčoňa* 'at Christmas').

Rarer temporal uses of non-separative inessive and contact case markers include: the future distance relation ('in X's time') encodable by the locative, or by the inessive or contact adpositions (e.g. in Welsh Romani, Austrian Sinti, Klenovec Rumungro, Florina Arli, and Rešitare); the telic extent relation ('in') encodable by the locative or by the inessive adposition (e.g. in Slovene and Florina Arli); and the atelic extent relation ('for') encodable by the contact adposition (in some Central dialects). Examples in (25)–(27) illustrate the various future distance metaphors:

(25) Welsh Romani (Sampson 1926: 179)
*Romerena pen kuši kūrken-**dī.***
get.married.3PL REFL.PL.ACC little week.PL-LOC
'They are getting married in a few weeks.'

(26) Austrian Sinti
*Me vau **an** trin čon.*
I come.1SG in three month
'I will come in three months.'

(27) Klenovec Rumungro
*Mēg na uštidīňom mīro paso, uštidā le **pe** čhoneste.*
still not got.1SG my passport get.1SG.FUT him.ACC on month.LOC
'I have not received my passport yet, I will get it in a month.'

The second major temporal metaphor concerns the extension of non-separative **adessive** markers to the anterior-durative relation ('until'). Encoding of the anterior-durative by the adessive adposition *ke* 'at, to' alone is attested in a single dialect, viz. in Lithuanian Romani; see example (28). Nevertheless, in numerous dialects, this adposition combines with limitative particles (e.g. *dži*, *pos*, *bis*);[13] see example (29). In the local domain, the complex expression *dži ke*, *dži te* (etc.) 'up to' encodes the limitative adessive, i.e. adessive localisation of directive orientation with an additional limitative feature. Importantly, this complex expression is found not only in dialects where there are complex adpositions (e.g. Prilep Arli and Sepečides; see Section 17.1), but also in dialects without complex adpositions (e.g. Russian Romani, Rumungro, Lovari,

and Taikon Kalderaš). This means that *ke* (*te*) in the limitative expression must be treated as a genuine adessive marker rather than as a general local adposition, and so we are dealing with an adessive metaphor. On the other hand, inessive encoding of the anterior-durative relation is very rare, being found only in a few Central dialects; see example (30).

(28) Lithuanian Romani
Kaj jov dživďa ***ke*** *lynaj?*
where he lived.3SG to summer
'Where did he live until the summer?'

(29) Klenovec Rumungro
Kāj bešla ***ži*** ***k****-o ňilaj?*
where sit.3SG.FUT LIM to-ART summer
'Where is he going to live until the summer?'

(30) Slovak Romani (Lučivná)
Avava kija tumende ***andr****-o sombat.*
be.1SG.FUT at you.PL.LOC into-ART Saturday
'I will be in your place until Saturday.'

The third major temporal metaphor is the use of the **separative adessive** case markers, the inflectional ablative and/or the adposition *katar ~ tar* 'from', for the posterior-durative temporal relation ('since'). The ablative is employed in Finnish Romani, the Northeastern dialects, most Central dialects, and in Taikon Kalderaš; the adposition is used in Šóka Rumungro, Arli of Kumanovo Prilep and Florina, Lovari, Kumanovo Gurbet, and Vălči Dol; see the examples in (31)–(32).

(31) Finnish Romani (Helsinki)
*Me passā joi hin jivutas dāri nījalesko čōnes-****ta.***
I believe.1SG s/he be.3 lived.3SG here summer.GEN month-ABL
'I think he has lived here since June.'

(32) Vălči Dol
Misljarav vov si kate ***katar*** *o juni.*
think.1SG he be.3 here from ART June
'I think he has been here since June.'

Only rarely are the separative adessive markers used to encode other temporal relations. First, a few dialects employ the ablative to encode simultaneous relations, with clock time in some Central dialects (e.g. *duje orendar* 'at two o'clock') and with parts of days in a few Balkan dialects (e.g. Sepečides *javinatar* 'in the morning'). Second, the ablative may encode past distance ('ago'), as attested in Florina Arli and Taikon Kalderaš; see example (33). Finally, the ablative or the adposition *katar* 'from' may be used in the atelic extent ('for') relation, as attested in some dialects of the Balkans (e.g. Malokonare, Muzikanta, Nange, Rakarengo, Rešitare, and Vălči Dol); see example (34).

(33) Florina Arli
*Duje bresen-**dar** prandindom me čaves.*
two.OBL years-ABL married.TR.1SG my.OBL son.ACC
'Two years ago I had my son married.'

(34) Malokonare
*Živoizava ando gav **tar** panč breš.*
live.1SG in.ART village from five year
'I have lived in the village for five years now.'

The fourth major temporal metaphor is the use of **horizontal** adpositions for sequence or distance temporal relations: the anterior adposition *angle* (*anglal*) 'in front of' is used to encode anterior sequence ('before') and/or past distance ('ago'), and the posterior adposition *pal* (*palal*) 'behind' is used to encode posterior sequence ('after') and/or future distance ('in X's time'). Unless borrowed adpositions are employed, this metaphor is almost universal in Romani (with rare exceptions in the future distance relation). The four functions are illustrated in (35). Horizontal adpositions appear to have no other temporal uses.

(35) Slovak Romani (Lučivná)
a. *Hin khere vareko **angl**-o dilos?*
be.3 at.home someone in.front.of-ART noon
'Is anyone at home before noon?' (anterior sequence)
b. *Ulĭle **angl**-o pandž berš.*
were.born.3PL in.front.of-ART five year
'They were born five years ago' (past distance.)
c. *Avava ke tu **pal** o dilos.*
come.1SG.FUT at you behind ART noon
'I will come to your place after noon.'

d. *Imar* ***pal*** *o berš avla miro.*
already behind ART year come.3SG.FUT mine
'[It] will be mine already in a year's time.'

Finally, temporal metaphors based on vertical, proximate, or peripheral localisations are unattested, with two rare exceptions. The Muzikanta dialect makes use of the medial adposition *maškar* 'between, among' in the simultaneous relation with parts of day and seasons (e.g. *maškarə zisəste* 'at day', *maškar evindiste* 'in the winter'). And, calquing Hungarian, some Central dialects and Lovari employ the inferior adposition *tel* (*telal*) 'under' in the future distance ('in X's time') and the telic extent ('in') relations; see the examples in (36).

(36) Šóka Rumungro
a. *Talākozinaha* ***tal*** *o pāndž dī.*
meet.1PL.FUT under ART five day
'We will meet in five days.' (future distance)
b. ***Tal*** *o pāndž dī āri sastʲja.*
under ART five day out recovered.3SG
'S/he recovered in five days.' (telic extent)

17.6. Internal diversity

Diversity asymmetries among the localisation values arise due to various developments: through the so-called ablative shift, through the development of complex adpositions, and especially due to numerous extensions. Ablative shift is most likely to affect horizontal localisations followed by vertical localisations, and it is least likely to occur with core and peripheral localisations (see Section 17.1). The development of complex adpositions is least likely to affect the core localisations, especially the adessive (see Section 17.1). On the other hand, extensions (see Section 17.4) and some other developments create greater cross-dialectal diversity in the peripheral and the core localisations. The generalised diversity hierarchies are shown in (37): (a) a hierarchy among the four localisation groups; (b) a hierarchy among peripheral localisations; (c) a hierarchy among core localisations; and (d) a hierarchy among axis localisations:

(37) a. Peripheral > core > proximate > axis
b. Peripheral: other > medial

c. Core: adessive > contact > inessive
d. Axis: superior > anterior > posterior > inferior

Among the **core** localisations, the adessive is the most diverse. First of all, there is the inherited variation between the adpositional variants *ke* and *te* 'at, to' in the non-separative orientation, as well as between the variants *katar* and *tar* 'from' in the separative orientation. This variation has led to cross-dialect diversity through option selection. In the non-separative, most dialects generalise *ke*, while Florina Arli and Muzikanti generalise *te*; both variants are retained in Welsh Romani and Nange. In the separative, most dialects generalise *katar* (*kathar*, *khatar*, *kata*, *kat*, *kotar*, *kote*), while modern Sofia Erli, Montana Kalajdži, Malokonare, and Kumanovo Gurbet generalise *tar* (*thar*, *atar*); both variants are retained in Arli of Gilan and Kumanovo and in Ajia Varvara. Second, the non-separative adposition *ke* 'at, to' shows innovative forms: *kije* or *kija* in the North Central dialects, and *kaj* or *ka* in some Vlax and Balkan *zis*-dialects (e.g. North Vlax, Xoraxane, Ajia Varvara, Rešitare, Kalburdžu, Varna Kalajdži, Gadžikano and Kaspičan).[14] Finally, specific case markers for directive adessives have been developed in numerous dialects: the adposition *karig* (*karing*, *kari*, *kori*, *koro* < **kaja-rig* 'this side') in Vlax and most Balkan dialects; the inflectional dative in various dialects; and more.

Separative core adpositions may be lost due to generalisation of the inflectional ablative. The loss may affect all core localisations (e.g. in Welsh, Estonian and Lithuanian Romani, Varna Bugurdži, and Nange), the inessive only (e.g. in Kalburdžu), or skip the inessive (e.g. in Polish Romani). On the other hand, inflectional core marking may be lost due to generalisation of the core adpositions (in various dialects). The contact localisation may be subject to ablative shift, while there is no ablative shift in the other core localisations. The original inessive and contact adpositions (of both orientations) may be replaced through adessive extension, and the non-separative adessive adposition *ke ~ te* may be replaced through proximate extension. The separative adposition *opral* 'from the surface of' can be also replaced through posterior extension (in some Central dialects).

In Kaspičan and Varna Gadžikano, the separative inessive adposition *andar* 'out of' is lost due to decomposition of localisation and orientation marking. While the non-separative inessive is encoded by the preposition *an* (< **andre*) with a locative case assignment to the object noun, the non-separative inessive is encoded by the same preposition with an ablative case assignment. Thus the preposition *an* 'in, into, out of' encodes the inessive localisation (without

cumulating the category of orientation), and the case assigned to the object noun encodes orientation; see examples in (38). Similarly, some dialects have decomposed localisation and orientation marking in the contact localisation (e.g. in Ajia Varvara, the orientation-indifferent contact marker is the adposition *pa* 'on, from the surface of', developed from the original separative adposition *opral*).

(38) Kaspičan
   a. *Odja sine* ***an*** *keres-**te***.
      that.F be.3 in/out.of house-LOC
      'She is in the house.'
   b. *Oj inkista* ***an*** *keres-**tar***.
      she got.out.3SG in/out.of house-ABL
      'She came out of the house.'

The **proximate** localisation is relatively stable, although the original adposition *paš* 'by, beside' is replaced by the limitative particle *dži* in some Balkan dialects (e.g. Velingrad Yerli, Varna Bugurdži, and Muzikanta). The localisation 'beside' can be differentiated from the proper proximate ('by') through borrowing (see Section 17.7).

Axis localisations are on the whole very stable, with the exception of the superior localisation. The inferior localisation can only be subject to ablative shift. The horizontal localisations may be, in addition, supplemented (but not replaced) by other indigenous local adpositions through extension, more likely in the anterior than in the posterior. The superior is the least stable axis localisation. In Early Romani, there was no distinction between the (non-separative) contact and the superior, both localisations being encoded by *opre* 'on; over'. While this conflation is retained in some dialects of the Balkans, numerous dialects have innovated the superior, replacing the original adposition through various extensions (e.g. Šóka Rumungro *uppe* 'on' vs. *perdal* 'over') or reinforcing it by ablative shift (e.g. Lučivná Slovak Romani *pre* 'on' vs. *upr-al* 'over'), comparative marking (e.g. Klenovec Rumungro *pre* 'on' vs. *upr-eder* 'over') or adverbs (e.g. Bohemian Romani *pre* 'on' vs. *upre pre* 'over', lit. 'above on').

Peripheral localisations are on the whole very unstable, being affected by numerous extensions of other local adpositions and case markers, both peripheral and non-peripheral. The medial localisation appears more stable than the other peripheral localisations: the original adposition *maškar* (*maškaral*) 'between, among' is retained in most dialects.

17.7. Borrowing

Adverbials encoding peripheral localisations are, on the whole, much more likely to be borrowed than non-peripheral adverbials, and proximate adverbials are more likely to be borrowed than core or axis adverbials. Several localisation hierarchies in borrowing of case markers (mostly adpositions) are shown in (39): (a) a hierarchy among the four localisation groups; (b) a hierarchy among peripheral localisations; (c) a hierarchy among non-separative core localisations, which partly differs from (d) the hierarchy among non-separative core localisations; and (e) a hierarchy among axis localisations. The borrowing hierarchy for local adverbs is given in (40). All hierarchies concern cross-dialectal frequency of borrowing, and do not necessarily engender implicational generalisations.

(39) a. Peripheral > proximate > core > axis
b. Peripheral: oppositive > perlative, translative, circumlative, sequentive > medial
c. Core (non-separative): adessive > inessive > contact
d. Core (separative): adessive > contact > inessive
e. Axis: vertical > horizontal

(40) Peripheral > lexical, proximate > core, axis

There are no proper loans of (non-separative) contact adpositions or horizontal adpositions. Lithuanian Romani is the only dialect to have a borrowed preposition in the anterior localisation (*pret'u*, *pretiv*, *protiv* 'in front of, opposite' from Russian). However, the source function of the preposition is oppositive, and the extension to the anterior function is an internal development. Similarly, the only borrowed preposition of (non-separative) contact, Welsh Romani *tap* 'on', is a result of internal grammaticalisation of the English noun *top*. There are no borrowed adpositions in posterior localisation.

Loan case markers are extremely rare in the (non-separative) inessive and the vertical localisations. Borrowed (non-separative) inessive case markers are well attested only with names of localities and especially countries that are themselves loans from the relevant contact languages: e.g. *v*(*əv*) *Rusija* 'in Russia' (< Bulgarian) in many dialects of Bulgaria, Kaspičan *Rusijada* 'in Russia' (< Turkish), Šóka Rumungro *Lenďelbe* 'in Poland' (< Hungarian).[15] A borrowed (non-separative) inessive adposition with common nouns is only attested in Prilep Arli (*vo* 'in' from Macedonian). Slovene Romani is the only

dialect that borrows vertical prepositions (*izpod* 'under' and *iznad* 'over' from Slovene). In all instances, the borrowed case markers alternate with indigenous ones.

Somewhat more frequent are loans of (non-separative) adessive and especially proximate adpositions. Well attested are borrowed adessive directives ('towards'): e.g. *gegn*, *nax* and the limitative *bis* 'up to' in Core Sinti (from German), *za* in Prilep Arli (from Macedonian), and the postposed particle *felē* 'towards' in Rumungro and Lovari (from Hungarian).[16] Proximate loans include *uze* 'by, beside' in western Rumungro, Prilep Arli and Dasikano (from Serbian/Croatian) and *konda* 'by' in Ajia Varvara (from Greek). There are also specific loans in the localisation 'beside': *nebn* and *langs* in Core Sinti (from German) and *zdravan* in Slovene Romani (from Slovene). Roman uses borrowed expressions both in the adessive and in the proximate localisations. However, the adessive *use* 'at, to' must have undergone an internal extension from proximate localisation (see Section 17.4), and the proximate *mere* 'by, beside' has probably developed from the place interrogative *mere* 'whither, which way' (< Hungarian *merre*).

Borrowed core adpositions of the separative orientation are relatively frequent. Yerli has completely replaced the indigenous separative adpositions of all core localisations by the preposition *is* 'from, out of' (from Bulgarian).[17] The loan functions as a general core separative preposition in the dialect; see examples in (41). In the modern Core Sinti dialects and in Roman, the preposition *fon* (*fun*, *fa*) 'from' (from German) completely replaces the indigenous separative adpositions in the adessive and contact localisations (see the examples in 42), while the indigenous inessive adposition (cf. Sinti *dran*, Roman *and*(*a*)*r* 'out of' < **andar*) is retained. Nevertheless, inessive uses of the borrowed *fon* are attested; see the Manuš example in (43).

(41) Yerli

a. *Oj iklisti* ***is*** *ko khər.*
she came.out.3SG from at house
'She came out of the house.' (inessive)

b. *Ləskiri phən irinəla* ***is*** *ko pazari.*
his sister turn.3SG from at market
'His sister returns from the market.' (adessive)

c. *Of hulejla o čantəs* ***is*** *ko masa.*
he take.down.3SG ART bags from at table
'He takes the bags down from the table.' (contact)

(42) Austrian Sinti
a. *Džas lo* ***fon*** *o kher je kota veg.*
went.3SG he from ART house one piece away
'He went away from the house.'
b. *Džias lo buter nit teli* ***fon*** *laki bukla.*
went.3SG he more not down from her back
'He didn't get off her back anymore.'

(43) Manuš (Valet 1991: 129)
Džijas li vri ***fun*** *o hole ruk.*
went.3SG she out from ART hollow tree
'She went out of a hollow tree.'

Borrowed separative adpositions with names of localities and/or countries are attested in some dialects of the Balkans (e.g. Prilep Arli *od Radoviš* 'from R', Varna Bugurdži *ot Varšava* 'from Warsaw', with Slavic *od*/*ot* 'from'). The Gadžikano dialect borrows Turkish ablative forms in this function (e.g. *Varšava-dan* 'from Warsaw', *Polša-dan* 'from Poland'). Klenovec Rumungro employs the preposition *mere* 'from' (< Hungarian *merre* 'whither, which way') in the adessive and contact localisations. As in Roman, its adpositional function has probably developed through internal change. Finally, some separative core adpositions are internally derived from loan adpositions, i.e. they are not loans themselves: e.g. Slovene Romani *uz-ar* (*uzal*, *zuro*) 'from, out of' from the Croatian loan *uzo* (*zu*) 'at, to'.

Among peripheral localisations, borrowed medial adpositions ('between, among') are the least frequent, being attested only in some dialects of Bulgaria (e.g. Yerli, Malokonare, Varna Kalajdži, Rešitare *meždu*, *meždi*, from Bulgarian). Circumlative ('around') and/or sequentive ('past, along') loans are more frequent. Most commonly they have been borrowed from current L2s (e.g. Lithuanian Romani *vakrug* from Russian; Slovene Romani, Prilep Arli, Sofia Erli and Muzikanta *okulo*, *okolu*, *okolo* from Slovene, Macedonian, and Bulgarian; Nange and Rešitare *kraj*, *pokraj* from Bulgarian), less commonly from recent L2s (e.g. Klenovec Rumungro *kerīl* from Hungarian). Perlative ('through') and/or translative ('across, over') loans are common, too. Current loans are found in German and Austrian Sinti (*durx* from German), West Slovak Romani (*ces* from Slovak), Slovene Romani (*čezo* from Slovene), and Kosovo Bugurdži (*preko* from Serbian). In some dialects, the prepositions are borrowed from an older L2: the German-derived *durx* in Hungarian Sinti,

Estonian and Russian Romani, and the Serbian/Croatian-derived *preke* (*preko*) in Slovene Romani and some Central dialects. In Klenovec Rumungro, the translative preposition *prēk-al* is an internal ablative derivation from the Serbian/Croatian-derived perlative preposition *prēke*.[18] Oppositive adpositions are borrowed most frequently, among peripheral localisations as well as in general. We find loans from current L2s (e.g. Core Sinti *gegn* from German, Lithuanian Romani *pret'u*, *pretiv*, *protiv* from Russian, Polish Romani *napšećiv* from Polish, West Slovak Romani *proci* from Slovak, Slovene Romani *nasproti* from Slovene, Bunkuleš Kalderaš and Priština Gurbet *protiv* from Serbian, Prilep Arli *sproti* and *karšia* from Macedonian, and Velingrad Yerli *sreštu* from Bulgarian) and recent L2s (e.g. Florina Arli *karšia* from Macedonian).[19]Among local adverbs, those of some peripheral localisations, especially the circumlative, are commonly borrowed (e.g. Lithuanian Romani *vakrug* 'around' from Russian; Taikon Kalderaš *řoata* and *krugom* 'around' from Rumanian and Russian, respectively; Sinti *langs* 'along' from German).[20] Loans of proximate adverbs are rarer (e.g. Šóka Rumungro *kezē* 'near' from Hungarian; Vălči Dol *blisko* 'near' from Bulgarian). On the other hand, there is no borrowing of adverbial word-forms in the core and axis localisations (for borrowing of orientation markers see Chapter 18). Since some dialects borrow lexical local adverbs, rather than deriving them from (borrowed) adjectives or nouns (e.g. Šóka Rumungro *jobra* 'to the right', *balra* 'to the left', *idegenbe* 'abroad', from Hungarian), the following borrowing hierarchy appears to hold for local adverbs: peripheral > lexical, proximate > axis, core.

Chapter 18
Orientation

The category of orientation is encoded in local expressions: local case markers (adpositions and synthetic cases), local adverbs, and local pro-words (interrogatives, indefinites in a wide sense, and deictics). There are three cross-cutting categories: localisation with case markers and adverbs, and lexicality and deictic distinctions (e.g. specificity) with pro-words. Importantly, the category of orientation is distinct from the category of localisation: while localisation encodes different spatial configurations of a figure object with regard to a ground object (see Chapter 17 for details), orientation specifies whether such spatial configurations are conceived of as actual ones, or ones that have been abandoned or ones to be assumed through movement of the figure object.[1] The values of the category of orientation vary in different dialects and for different structures within a dialect. In other words, there are different orientation paradigms. Orientation values may be defined as overt distinctions on a cognitive map; different orientation paradigms conflate different functions on the map. The functions relevant for Romani are defined by two semantic features: the static vs. dynamic character of the event associated with the local expression and, provided the event is dynamic, the sort of local anchoring of the event with regard to a reference point. For example, local pro-words in Šóka Rumungro encode four distinct orientation values. Consider the following examples with local interrogatives:

(1) *Kāj bešes?*
where sit.2SG
'Where do you live?'

(2) *Kija džas?*
whither go.2SG
'Where are you going?'

(3) *Kēre džaha?*
which.way go.2SG.FUT
'Which way will you go?'

(4) *Kathar džaha?*
which.way go.2SG.FUT
'Which way will you go?'

(5) *Kathar aves?*
whence come.2SG
'Where are you coming from?'

The *stative* interrogative *kāj* 'where' in (1) is associated with a static event. The *directive* interrogative *kija* 'whither' in (2) is associated with a dynamic event and indicates movement towards a reference point. The *perlative* interrogative *mēre* 'which way, through where' in (3) is associated with a dynamic event and indicates movement through a reference point (environment). The interrogative *kathar* 'whence; which way, through where' may have the perlative function as in (4), but it also has a *separative* function in (5) where it is associated with a dynamic event and indicates movement away from a reference point. Thus, local pro-words in Šóka Rumungro encode four orientation values: stative, directive, perlative, and separative-perlative. The function 'perlative' may be encoded by two sets of expressions (e.g. the interrogatives *mēre* and *kathar*), and the function 'separative' is always conflated with the function 'perlative'.

Local adverbs in Šóka Rumungro encode different orientation values: apart from the stative and the directive, there is also a *static separative* and a *dynamic separative* in some localisations; and there is no encoding of the perlative orientation. Examples (6)–(7) illustrate the static and dynamic separatives, respectively. The adverb *dūral* in (6) is associated with a static event, while the adverb *dūraltū* in (7) is associated with a dynamic event (encoded by a verb of movement).

(6) *Dūral na dikhav.*
far.STC.SEP NEG see.1SG
'I cannot see from a distance.'

(7) *Ājom dūraltū.*
come.PRET.1SG far.DYN.SEP
'I came from far.'

Our investigation of the category of orientation is impeded by lack of data for some of the less salient orientations. An overt distinction between the static

and the dynamic separative is only described for Šóka Rumungro, and we have little data on the perlative orientation. Mostly, then, we have investigated asymmetries between the basic orientations: the stative, the directive, and the separative. The separative stands out according to all criteria: while on the one hand it is most likely to extend and tends to be the most complex, most exposed, and most borrowable value, it is on the other hand the least differentiated value and the value that is least prone to internal renewal. The stative is the least likely value to be exposed and borrowed, while the directive is the least likely value to be extended, and the most likely value to be renewed through internal developments. The mutual position of the stative and the directive is thus ambiguous. There are two linear orderings of the three basic orientations:

(8) Separative–stative–directive
(9) Separative–directive–stative

The ordering in (8) is relevant for the criteria of extension and diversity due to internal renewal, while the ordering in (9) is relevant for the criteria of exponence and borrowing. The criteria of complexity and differentiation do not render any asymmetry between the stative and the directive. Whenever we have access to the data, the position of the perlative orientation is on an extreme of a scale: the perlative is the most borrowable and the least differentiated value.

18.1. Extension

While ablative forms were, in Early Romani, associated with the separative orientation, in many dialects we find ablative forms in the stative and/or directive orientations as well. This is frequently the case with adverbs and adpositions, but never the case with pro-words or synthetic case markers. With adverbs, the ablative forms in *-al* extend from the separative orientation to the stative and/or directive orientations (e.g. *pal-al* ‘in/to the back’ < ‘from the back’). There were probably no specific adpositions for the separative orientation of most localisations in Early Romani (see Chapter 17). The stative/directive ablative adpositions found in some dialects probably arose directly through grammaticalisation of ablative adverbs (e.g. *pal-al* ‘behind’ < ‘from the back’), rather than through an extension of separative adpositions into the stative/directive orientation. For the sake of convenience, we use the term *ablative extension* to refer to both processes.

Ablative extension is sensitive to the category of localisation, i.e. adverbs and adpositions of different localisations may show a different degree of ablative extension (see Chapter 17). Here, we will illustrate ablative extension with expressions of anterior localisation. The separative anterior relation ('from the front of') is rarely encoded as a case relation. If it is, then the ablative adposition *angl-al* is used (10). Separative anterior adverbs ('from the front') are always ablative in form, either reflexes of *angl-al* or derivations thereof (11). Both examples are from Šóka Rumungro.

(10) *Dža anglal mre jakha.*
go from.the.front.of my.PL eye.PL
'Get out of my sight.' (lit. 'Go from the front of my eyes')

(11) *Ājom ānglaltū.*
come.PRET.1SG from.the.front
'I came from the front.'

Dialects differ, however, in the extent of the use of ablative forms in the stative and directive orientations, viz. in stative/directive adpositions ('in front of, to the front of'), stative adverbs ('in the front'), and directive adverbs (to the front'). Table 18.1 charts the various attested patterns; the plus sign indicates the use of an ablative form.

Type F is found, for example, in Welsh Romani, Ajia Varvara, and Dasikano, and Type I in Core Sinti, Gadžikano, Muzikanta, and Malokonare. An import-

Table 18.1. Distribution of ablative forms of adpositions and adverbs

	Adposition	Adverb	
	Stative/directive	Stative	Directive
Type A: Estonian, Slovene R	–	–	–
Type B: Roman	–	±	–
Type C: East Slovak R	–	+	±
Type D: Rumungro	–	+	+
Type E: Slovak R (Zips)	±	+	+
Type F: see below	+	–	–
Type G: Prilep Arli	+	±	–
Type H: Taikon Kalderaš	+	+	–
Type I: see below	+	+	+

ant generalisation over the patterns is that an ablative form of the directive adverb implies an ablative form of the stative adverb (of the same localisation), but not vice versa. In diachronic terms this means that ablative extension in adverbs proceeds from the separative orientation to the stative orientation, and only from the stative orientation to the directive orientation (e.g. 'from the front' > 'in the front' > 'to the front'). With adpositions, the ablative extension proceeds from the separative orientation to both other orientations simultaneously, as there is mostly no distinction between the stative and the directive orientations. There is no implicational relation between the presence of ablative forms in adpositions and in adverbs. For example, in Rumungro (Type D), there is a non-ablative preposition *angle* and an ablative stative/directive adverb *āngl-al*, while in Ajia Varvara (Type F), there is an ablative preposition *angla* (< **angl-al*) and a non-ablative stative/directive adverb *angle*.

18.2. Exposition

Early Romani encoded two orientation values: the stative-directive (not distinguished in congruence with the Balkan languages) and the separative. The conflation of the stative and the directive, and hence a greater exponence of the separative, has been retained as the most frequent pattern: it is almost the rule in those case markers that encode orientation (see Chapter 17), and frequent in adverbs and pro-words. The extension of ablative adverbs to the stative orientation (see Section 18.1) has created a conflation of the separative with the stative, and hence a greater exponence of the directive. This exponence pattern is restricted to adverbs (of some or all localisations) in some dialects. A greater exponence of the stative is restricted to local pro-words, and is cross-dialectally rare (e.g. the stative *kaj* 'where' vs. the directive-separative *kā-rīg* 'whither; from where' in Latvian Romani). To sum up, all three basic orientation values may show a greater exponence than the other values. Nevertheless, differing cross-dialectal frequency of these patterns indicates that the separative is more likely to be exposed than the directive, which is in turn more likely to be exposed than the stative.

18.3. Internal diversity and borrowing

There is a partial mismatch between the results of the criterion of cross-dialectal diversity due to internal renewal and the criterion of borrowing. The

separative orientation is, on the one hand, the most resistent to internal renewal (clearly with interrogatives, less so with deictics, case markers, and adverbs), while, on the other hand, it is relatively susceptible to borrowing (especially with orientation markers in deictics, and with case markers). There appears to be a slight tendency for the directive orientation to be more prone to internal renewal than the stative orientation (with interrogatives and deictics, and possibly with adverbs). Directive markers are also more likely to be borrowed than stative markers (with deictics). As far as borrowing of pro-words is concerned, the perlative orientation appears to be the most susceptible to borrowing. Unfortunately, there are gaps in our data on adverbs, and we have little information on the perlative orientation. Consequently, the relevant generalisations must be considered preliminary.

The separative orientation shows the least diversity with interrogatives and indefinites and, as far as internal renewal is concerned, also with deictics. Most dialects retain the interrogative *ka-tar* (*khatar*, *kathar*, *katyr*) 'from where' and deictics in one of the indigenous ablative suffixes (e.g. Rumungro *odo-thar* or Nange *odok-ar* 'from there'). The separative orientation may be renewed through grammaticalisation of the noun *rig* 'side'. This is rare with the interrogative (e.g. Latvian Romani *kā-rīg* 'from where; whither'), and somewhat more frequent with deictics (e.g. Polish Romani *do-ryk*, or East Slovak Romani *oda-rig* 'from there'). Separative deictics may also contain a combination of the grammaticalised noun and an ablative suffix (e.g. Bunkuleš Kalderaš *odo-ring-al* or Finnish Romani *too-ri-ta* 'from there').

There is more diversity in the stative and/or directive pro-words. Although many dialects retain the indigenous interrogative *kaj* 'where' and some of the indigenous deictics in *-Vj* or *-e* (e.g. *okoj*, *oke*, *orde* 'there'), there are also numerous innovative forms. In many dialects, stative and/or directive pro-words contain reflexes of the regular locative suffix *-te*, thus being parallel formations to the separative pro-words in the regular ablative suffix *-tar*. The interrogative *ka-te* 'where' is found in many dialects of Bulgaria (e.g. Sofia Erli, Varna Bugurdži, Muzikanta, Nange, Malokonare, Kaspičan, Gadžikano, Lom, and Varna and Montana Kalajdži) and in Crimean Romani. Deictics of this type (e.g. *ko-te* 'there') are much more widespread: they are found in the Sinti, Slovene Romani, Balkan and Vlax dialects and also in Latvian Romani; they are missing in Welsh and Finnish Romani, the Central dialects, and unattested in most Northeastern dialects.[2] A later reinforcement by the deictic root *-k-* in some Balkan and Vlax dialects has obscured the shape of these forms (e.g. Kosovo Bugurdži *ko-t-ka* < **ko-te* and *-k-*). Another common type of renewal is the grammaticalisation of the noun *rig* 'side', as already encoun-

tered with separative pro-words. Both the interrogative *ka-rig* 'where' (*kariga*, *karing*, *qari* etc.) and deictics of this type (e.g. *odorig* 'there') are found in some Northeastern dialects and some Balkan dialects (e.g. Gadžikano). Some other Balkan and Vlax dialects possess either only the interrogative (e.g. Varna Bugurdži, Nange, Iranian Romani, Austrian Lovari), or only the deictics (e.g. Arli of Gilan and Florina, Sepečides, Sofia Erli, Yerli, Taikon Kalderaš). The deictics (e.g. *tōri* < **odoja-rig*) are also found in Finnish Romani. Apart from these common innovations, there are a number of stative and/or directive pro-word formations that are restricted to individual dialects.[3]In some dialects, innovative interrogatives encode both the stative and the directive orientations, either completely replacing the indigenous interrogative (e.g. *kate* in Montana Kalajdži, *kati* in Muzikanta, and *qari*(*k*) in Iranian Romani), or being used as free variants (e.g. *kaj* or *kate* in Varna Kalajdži). More frequently, however, innovative forms are specialised in dialects that have created a distinction between the stative and the directive orientations, partly due to structural convergence with languages that possess such a distinction. There are three patterns. First, the indigenous *kaj* is retained in the stative orientation, while the directive is innovative (e.g. Latvian and Russian Romani *karik*, western Rumungro *kija*, and possibly Slovene Romani *kev*). Second, more rarely, the indigenous *kaj* is retained in the directive orientation, while the stative is innovative (e.g. Kaspičan *kate*). And, third, innovative forms are used in both orientations (e.g. Nange *kate* 'where' vs. *karig* 'whither'). The indigenous interrogative may be retained as a stative variant (e.g. Varna Bugurdži *kati* or *kaj* 'where' vs. *kariga* 'whither'). If the stative and the directive orientations are distinguished in deictics, the directive forms tend to be more innovative (e.g. Russian Romani *odoj* 'there' vs. *odorik* 'thither', or western Rumungro *odoj* vs. *onďa*).

Deictics (but not interrogatives) may contain borrowed orientation markers. Separative markers are attested in Sofia Erli (e.g. *iz-akatar* 'from here', with *iz-* from Bulgarian), Piedmontese Sinti (cf. *da-kaj* 'from here' with *da-* from Italian), and Austrian Sinti (e.g. *fon koti* 'from there'). Austrian Sinti also borrows directive markers (e.g. *kaj her* 'hither' and *koj hin* 'thither', with *her* and *hin* from German). No borrowed stative markers are attested.

There is also borrowing of whole pro-words. Borrowing of local interrogatives and especially deictics is extremely rare. In Šóka Rumungro, only perlative interrogatives are borrowed (cf. *mēre* 'through where', *ēre* 'through here', and *āra* 'through there' from Hungarian). In Roman, the Hungarian-derived interrogative *mere* is used (alongside the indigenous *kaj*) in the stative/directive orientation, while in the separative orientation, there is the form *mer-al*

(alongside the indigenous *katar*), which is an internal ablative derivation rather a direct loan. Borrowing of local indefinites is more frequent. There are attestations of stative/directive loans (e.g. South Slavic *nigde*, *nindźe*, *nikade* 'nowhere' in various dialects of the Balkans, Greek *kapu* 'somewhere' in Ajia Varvara, German-derived *iberol* 'everywhere' in Austrian Sinti) as well as separative loans (e.g. Bulgarian *otnjakade* 'from somewhere' in Sliven dialects). Šóka Rumungro borrows only perlative indefinites (e.g. *valamēre* 'through somewhere' from Hungarian), while indefinites of the other orientations are derived from indigenous interrogatives (e.g. *vala-kāj* 'somewhere', *vala-kija* 'to somewhere', *vala-kathar* 'from somewhere'). Slovene Romani borrows the stative indefinite *nigdi* 'nowhere' from Slovene, while the directive indefinite *ni-kev* 'to nowhere' is an internal de-interrogative derivation.

In those localisations that encode orientation, stative/directive case markers appear to be slightly more prone to internal renewal than the corresponding separative case markers. On the other hand, separative case markers are more likely to be borrowed than the corresponding stative/directive case markers. See Chapter 17 for details.

In adverbs, all orientations may be renewed through grammaticalisation of various function words. Thus in Šóka Rumungro, stative adverbs are derived by the morpheme *onď-* from the corresponding directive adverbs in some localisations (e.g. *onď-ānde* 'inside STA' < *ānde* 'inside DIR'). This stative marker is clearly related to the directive local deictic *onďa* 'thither'.[4] In Hungarian Lovari, the preposition *pe* 'on' serves as a directive marker (e.g. *pe opral* 'upwards'), and the ablative particle *tar* serves as a separative marker (e.g. *opral tar* 'from up'). The most frequent source of internal renewal, however, is the ablative extension, which is more likely to affect the stative than the directive orientation (see Section 18.1). If there is ablative extension, separative adverbs may be secondarily distinguished either through borrowed separative markers (see below), or through a reduplication of the original ablative suffix *-al* (e.g. *opr-al-al* 'from up' vs. *opr-al* 'up STA' in some varieties of East Slovak Romani).

Adverbial orientation markers are borrowable in all orientations, with conflicting asymmetries in different dialects. This may be connected to the availability and morphosyntactic transparency of orientation markers in the source languages. Some dialects borrow only separative markers: for instance, Šóka Rumungro *-tū* (e.g. *āvral-tū* 'from outside') from Hungarian, or Core Sinti *fon* (e.g. *fon vrial* 'from outside') from German. Other dialects borrow only stative/directive markers (e.g. East South Slavic *na-* in Gilan Arli, Sofia Erli, and Varna Kalajdži). Yerli borrows the separative marker *iz-* as well as the

stative/directive marker *na-* from Bulgarian (e.g. *iz-avrel* 'from outside', *na-avri* 'outside, out'). In Finnish Romani, only the directive marker *päi* is borrowed, from Finnish (e.g. *avri päi* 'out DIR'). Borrowing of whole word-forms of local adverbs is only attested with directive adverbs in our data (e.g. Erli *napered* 'to the front').

18.4. Complexity

The criterion of complexity does not seem to render any generally valid asymmetry. While in Early Romani, the separative orientation was consistently more complex than the stative/directive orientation (in pro-words, case markers, and adverbs), various developments have disturbed this asymmetry. Individual dialects show numerous patterns which may differ according to the structure involved, and there does not seem to be a simple way to generalise over the dialect-specific patterns.

In Early Romani, separative pro-words were more complex than corresponding stative and/or directive pro-words. Dialects that have created the stative/directive forms in *-te* have reduced this asymmetry (e.g. Varna Kalajdži *o-the* 'here, hither' vs. *o-thar* 'from here'). In a few dialects, directive pro-words are as complex as, or more complex than, separative pro-words (e.g. Latvian Romani *dārīg* 'hither; from here', or Taikon Kalderaš *ka-ring-ar* 'hither' vs. *ka-tar* 'from here'). Stative pro-words are only rarely more complex than directive pro-words (e.g. Kaspičan *ka-te* 'where' vs. *kaj* 'whither').

There is no clear complexity asymmetry with those case markers that encode orientation (see Chapter 17). There are dialects where stative/directive case markers are consistently less complex than the corresponding separative case markers (e.g. Yerli *ke* 'in, into, at, to, on' vs. *is ke* 'out of, from'), dialects where the opposite is the case (e.g. Lithuanian Romani *de* 'in, into', *ke* 'at, to', and *pe* 'on' vs. the synthetic ablative in the separative meanings 'out of, from'), as well as dialects where the stative/directive is less complex in some localisations but more complex in others (e.g. Kalburdžu *ka* 'at, to' vs. *ka-tar* 'from', but *ande* 'in, into' vs. the synthetic ablative in the separative meaning 'out of').

Separative adverbs are more complex than the stative/directive adverbs in those dialects that retain the Early Romani pattern. The fact that the ablative extension proceeds first to stative adverbs and only then to directive adverbs (see Section 18.1) means that the former are more likely to be more complex than the latter. Nevertheless, we have seen in Section 18.3 that adverbs of all

orientations are reinforceable through grammaticalisation or borrowing, and so there does not seem to be any absolute ban on increase of complexity in the stative and/or directive adverbs.

18.5. Differentiation

The separative is the least differentiated (of the three basic orientation functions) in terms of the cross-cutting category of localisation in case markers and adverbs, as it is restricted to some localisations only (see Chapter 17). At least in some dialects, the perlative orientation is the least differentiated in terms of deictic distinctions in deictic pro-words. Table 18.2 shows the local deictics in Šóka Rumungro, with four terms in the directive, the stative, and the separative(-perlative), but only two terms in the perlative proper.

Table 18.2. Local deictics in Rumungro

Deictic	Directive	Stative	Separative	Perlative
'here'	*anďa*	*adaj*	*āthar* (*adathar*)	*ēre*
'just here'	*akija*	*ākaj*	*akathar*	
'there'	*onďa*	*odoj*	*ōthar* (*odothar*)	*āra*
'just there'	(*am*)*okija*	(*am*)*okoj*	(*am*)*okothar*	

Chapter 19
Indefiniteness

The category of indefiniteness is encoded in indefinite pro-words, or indefinites. In a cross-linguistic study of indefinites,[1] Haspelmath (1997b) has isolated nine indefiniteness functions, and projected the semantic contingencies among these functions onto a universal semantic map (Figure 19.1). The functions are: specific known (Sk), specific unknown (Su), irrealis non-specific (I), conditional (Cn), comparative (Cm), question (Q), indirect negation (Ni), direct negation (Nd), and free-choice (F). We also include pro-words that function as universal quantifiers in our discussion, and consider 'universal' (U) to be a further indefiniteness function (as reflected in Figure 19.1).

Individual languages draw their own distinctions on the semantic map. These language-specific distinctions are the indefiniteness values of the language. Each indefiniteness value is encoded by a series of indefinites of different ontological values. The ontological category is the major cross-cutting category for indefiniteness. We use the following terminology of language-specific indefiniteness values: a *specific* series comprises the specific unknown function; a *negative* series comprises the direct negation function; a *free-choice* series comprises the free-choice function but not the universal function; a *universal* series comprises the universal function; and a *negative-polarity* series comprises the question and/or the conditional functions.

Three indefiniteness series may be reconstructed for Early Romani: a specific-to-negative series, a free-choice series, and a universal series (see Chapter 5 for details). The wide range of the specific-to-negative series, viz. from

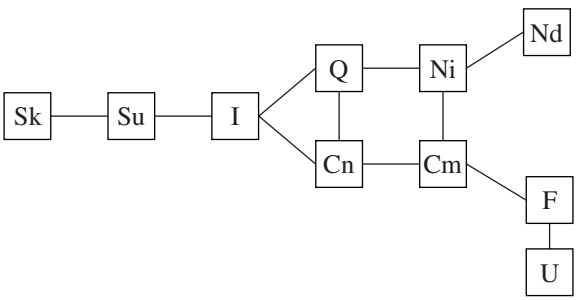


Figure 19.1. Semantic map of indefiniteness functions

specific via irrealis and question to negation, was probably inherited from Indo-Aryan and supported by Romani's Asian contact languages (see Elšík 2000a). The majority of current dialects, however, possess four series of indefinites: specific, negative, free-choice, and universal. The major difference with respect to Early Romani is that a distinct negative series has been created due to convergence with European languages. A typical example of the majority pattern is found in Central Slovak Romani as shown in Figure 19.2 (cf. the specific *vare*-series, the negative *ňi*-series, the free-choice *makar*-series, and the universal *sa*-series). Dialects, of course, differ in details of coverage of the individual indefiniteness values. For example in Welsh Romani, unlike in Central Slovak Romani, the negative series comprises the question and indirect negation functions as well as direct negation (see also Section 19.2). There are minority patterns as well. First, a few dialects retain the Early Romani situation in that they do not have a distinct negative series (see Section 19.2). Second, some dialects do not seem to have a distinct free-choice series, with the free-choice function being encoded by indefinites of the universal series. And third, borrowing of indefiniteness markers and words introduces further series from the source languages: for example, Xoraxane has developed a negative-polarity series through borrowing of the negative-polarity marker *i-* from Serbian/Croatian (see Section 19.5).

There is no single hierarchy among the major indefiniteness values. Free-choice indefinites are the most complex, the most diverse, likely to extend and may show extracategorial distribution. Free-choice markers are the most likely to be borrowed, while free-choice words are the least likely to be borrowed. Universal indefinites are the least complex, show medium diversity, and only rarely extend. Universal markers are the least likely to be borrowed, while

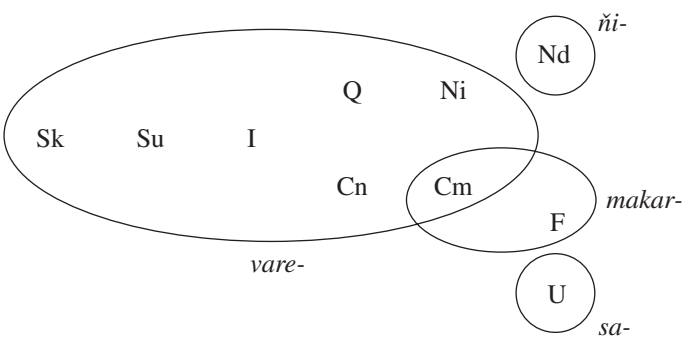


Figure 19.2. Indefiniteness marking in Central Slovak Romani

universal words are borrowed frequently. Negative indefinites are relatively complex, show medium diversity, do not extend, and may show extracategorial distribution. Both negative markers and negative words are likely to be borrowed. Specific indefinites show medium complexity and diversity and little extension. Both specific markers and specific words show medium susceptability to borrowing. The criterion of differentiation does not seem to render any indefiniteness asymmetry.

19.1. Complexity

The criterion of complexity renders the following asymmetry: free-choice > negative > specific > universal. Greater complexity correlates with greater transparency of indefiniteness marking. The older the marker, the less transparent it tends to be, and the less complex the indefinite is. Thus, the above complexity asymmetry partly derives from innovation asymmetries among different indefiniteness values (see Sections 19.4–19.5). For example, free-choice indefinites are frequently attested in incipient stages of grammaticalisation and free-choice markers are most likely to be borrowed, which renders free-choice marking the most complex in synchronic terms. Universal indefinites rarely contain a transparent marker, and so they tend to be the least complex. Free-choice indefinites may be formed by reduplication of interrogatives (see Section 19.4), which is rarely the case with the other indefinites. In some dialects, question indefiniteness is the least complex.

Indefinites derived from indefinites are more complex than de-interrogative indefinites. Of course, they are also more complex than the indefinites they are based on. De-indefinite indefinites have either the free-choice function (e.g. Taikon Kalderaš *voare-so-godi* 'whatever' < *voare-so* 'something',[2] Austrian Sinti *irgend-čomuni* 'anybody, whoever' < *čomuni* 'something'), or the negative function (e.g. Florina Arli *hič-čumuni* 'nothing' < *čumuni* 'something', Malokonare *ni-kacinende* 'nowhere' < *kacinende* 'somewhere'). In both cases, the base indefinites are specific, and so the specific value tends to be less complex. If specific indefinites contain two indefiniteness markers, they both mark the same indefiniteness value (e.g. Estonian Romani *vari-so-ta* 'something' alongside *so-ta* 'something'), i.e. the double-marked specific indefinites are not derived from indefinites of another indefiniteness value.

A few Vlax dialects have selected among variants of indigenous thing indefinites to distinguish negative functions from the other functions of the original specific-to-negative range. Thus in Varna Kalajdži, *khači* 'something,

anything' (< **kaj-či*) now covers specific, irrealis and negative polarity functions, while the more complex *khanči* 'nothing' (< **kaj-ni-či*) covers direct and indirect negation.[3] Similarly in Bunkuleš Kalderaš and Ajia Varvara, there is a distinction between the related *kha(j)ši* 'something, anything' and *khanči(k)* 'nothing'. Some dialects allow the use of interrogatives instead of indefinites in questions. Varna Kalajdži makes use of the least complex thing indefinite *či* in this function.

19.2. Extension

There are three kinds of extension in indefinites. First, universal indefinites are frequently used in the free-choice function as well, and some dialects lack a distinct free-choice value altogether. On the other hand, the free-choice-to-universal extension is also attested: in western Rumungro, the determiner *so-godi* 'all' has developed from the thing free-choice indefinite 'whatever' (see Section 19.5 for the suffix *-godi*). Next, free-choice indefinites are attested as extending to specific functions. Finally, specific-to-negative indefinites may develop into non-negative indefinites through replacement, and into negative indefinites through gradual loss of functions.

The shift from the free-choice function to specific functions has occurred with indefinites containing the prefix *vare-* (*voare-*, *var-*, *ver-*). The prefix is a loan of the Rumanian free-choice *(v)oare-* and is attested in this function in a few peripheral dialects (see Section 19.5). However, in most dialects where it occurs (e.g. the Northeastern and the North Central dialects, Catalonian Romani, modern Sofia Erli, Crimean Romani, the North Vlax, some South Vlax dialects, Ukrainian Vlax), it now has a specific function.[4] A similar functional shift must be assumed for the Prizren form *čhi-gode* 'something' and, according to one etymology (Elšík 2000c), also for the indefinites in *-moni* in Welsh Romani, the Northwestern dialects, Abruzzian Romani, Florina Arli, and Yerli.

The *kaj*-series of indigenous indefinites (see also Chapter 5) may be reconstructed as covering a wide range of indefiniteness meanings, probably from specific via irrealis and negative polarity to negative proper. We term such indefinites specific-to-negative. The Early Romani situation is best retained in older Finnish Romani as well as in the modern variety of Helsinki, and in some South Vlax dialects (e.g. Priština Gurbet, Macedonian Gurbet, Varna Kalajdži, Rešitare, and Vălči Dol). In Finnish Romani, the thing indefinite *či* and the indefinites based on the determiner **kaj* (viz. the determiner-person indefinite *ček*, the place indefinite *čēni*, and the time indefinite *čekkar*) are

specific-to-negative indefinites. In the above Vlax dialects, this holds for the person indefinite *khonik*, the thing indefinite *khanči*, and the place indefinite *katinende*. The specific-to-negative indefinites are accompanied by clause negation in negative contexts (e.g. Finnish Romani ***ma*** *čēr či* 'do not do anything'). Nevertheless, in most dialects that retain (some of) the indigenous indefinites, their function has been restricted by two kinds of developments. (A third type of development in the function of the indigenous indefinites, viz. selection among variants, has been discussed in Section 19.1).

First, the indigenous indefinites have been restricted to specific, irrealis and negative polarity functions through development of distinct negative indefinites. In Slovene Romani, the creation of the de-interrogative negative *ni-kon* 'nobody' has ousted the person indefinite *koniko* 'somebody' out of negative contexts. Similarly, Yerli *či* and Kosovo Bugurdži *hajči* 'something' are now used as non-negative indefinites, as there is the negative loan *ništo* 'nothing'. In Malokonare, the indigenous place indefinite has been prefixed with the negative marker *ni-* in negative contexts, and so now it is restricted to non-negative meanings (cf. *kacinende* 'somewhere, anywhere' vs. *nikacinende* 'nowhere').

Second, the indigenous indefinites (e.g. the thing *č(h)i*, the determiner/person *kek* or *tek*, and the forms based on the determiner) have gradually lost their specific and irrealis functions in western dialects. Thus in Welsh Romani and Kuopio Finnish Romani, they are negative polarity and negative indefinites, and in Core Sinti, some modern Finnish Romani varieties, and Catalonian, Polish, and Bohemian Romani, they are restricted to the direct negation function.[5] In Welsh Romani and Core Sinti, they may be even used without a clause negator and still interpreted as negative (e.g. Hungarian Sinti *me dikjom či* 'I saw nothing'). The new specific forms have mostly developed from free-choice indefinites (e.g. Bohemian Romani *či* 'nothing' < *'something, anything, nothing' vs. *vare-so* 'something' < *'anything whatsoever'; see above). The gradual loss of non-negative functions in the indigenous indefinites appears to be due to inherent functional shift, rather than due to replacement (as in the first type of development).

19.3. Extracategorial distribution

Free-choice indefinites may show extracategorial distribution in that they are used as connectors (e.g. Bunkuleš Kalderaš *sargod* 'as soon as' < 'in any manner'). Negative indefinites may develop into negators (e.g. North Vlax *či* 'not'

< *'nothing', or Welsh Romani *kek* 'not' < 'none, no'). There appears to be no extracategorial use of specific or universal indefinites.

19.4. Internal diversity

While free-choice marking shows the greatest internal diversity, there does not seem to be any obvious asymmetry among the other indefiniteness values. We consider indefinites developed from internal resources as well as internal grammaticalisation of borrowed elements.

As for free-choice indefinites, some dialects construct them by morphological reduplication of interrogatives of the corresponding ontological category (e.g. Ajia Varvara *kon-kon* 'whoever' < 'who-who'). In Latvian Romani, a negator is used to connect the interrogatives (e.g. *kon-na-kon* 'whoever' < 'who-not-who'). Free-choice markers may derive from subjunctive constructions containing the verb 'be' (e.g. Sepečides *so ti si* 'whatever' < 'what it may be'). Some East Slovak varieties have developed the prefix *mijel-* (e.g. *mijel-ko* 'whoever') through grammaticalisation of *mi jel* 'let him/her be', consisting of the optative particle *mi* (< **muk* 'let, leave') and a contracted 3SG subjunctive form *av-el* of the verb 'be'. There are also constructions based on the verb 'want' (e.g. Sepečides *so mangesa* 'whatever' < 'what you.SG want'), with a greater degree of grammaticalisation in some dialects (e.g. East Slovak Romani *ko-kam* 'whoever' with *-kam* < **kames* 'you.SG want'). Kosovo Bugurdži has the free-choice marker *kudžanla-* (< *ko džanla* 'who knows'). The free-choice prefix *fer-* in Taikon Kalderaš (e.g. *fer-savo* 'any whatsover') is, in all likelihood, an internal grammaticalisation of the Rumanian-derived focus particle *feri* 'only'. All these sources of free-choice marking are well attested cross-linguistically, and so they may be independent innovations in Romani. However, in some cases, structural convergence is more likely: for example, Sofia Erli *savo te ovel* 'any whatsoever; what he/it may be like' probably calques Bulgarian *kakăv da e* of the same function and structure.

Internal innovation is rarer in the other indefiniteness values.[6] Ukrainian Romani dialects have grammaticalised the local interrogatives *t'eu* or *kaj* 'where' as specific markers (e.g. *t'eu-ko* or *kaj-ko* 'somebody'), due to convergence with Ukrainian, where the specific prefix *de-* derives from the local interrogative.[7] The negative determiner *či-jek* 'no, none' in some North Vlax dialects contains the negator *či* 'not'. The negative prefix *gwar-* in Hungarian Sinti (e.g. *gwar-či* 'nothing') is a result of internal grammaticalisation of the German-derived negative particle *gwar* 'not, not at all'. Some dialects have re-

analysed the borrowed determiner/person indefinite *sako* 'every; everybody' (see Section 19.5) as containing the person interrogative *ko*(*n*) and an indefiniteness prefix *sa-*, which also exists in many dialects as a universal particle (of various ontological values). Thus in Prizren Arli, there is the de-interrogative place indefinite *sa-kote* 'everywhere', the determiner *sa-kova* (masculine) ~ *sa-koja* (feminine) 'every', which is formed as if derived from a demonstrative, and the nominalisation *sa-ben* 'everybody', which serves as a person indefinite. Šóka Rumungro has created the de-interrogative place indefinites *sa-kāj* 'everywhere', *sa-kija* 'to everywhere', and *sa-kathar* 'from everywhere', and Slovene Romani has created the de-interrogative person indefinite *sa-kon* 'everybody'.[8]

19.5. Borrowing

The criterion of borrowing renders two partly conflicting asymmetries, depending on whether one considers borrowing of indefiniteness markers (1) or borrowing of whole indefinite word-forms (2):

(1) free-choice > negative > specific > universal
(2) negative, universal > specific > free-choice

On both hierarchies, the value 'negative indefiniteness' is more prone to borrowing than the value 'specific indefiniteness'. This can be formulated in implicational terms: if the value 'specific' is borrowed, then the value 'negative' is also borrowed (specific → negative). There is one significant type of exception to this statement, viz. borrowed determiners, which we shall deal with below. The position of the value 'free-choice indefiniteness' differs on hierarchies (1) and (2); in fact, the position on one hierarchy is the exact opposite of the value's position on the other: free-choice markers are the most likely to be borrowed among the markers, whereas free-choice indefinites are the least likely to be borrowed among the indefinite word-forms. We find similar results with the value 'universal indefiniteness': in (1) we see that universal markers are not likely to be borrowed, while in (2) we see that universal indefinite word-forms are very likely to be borrowed. There is no implicational hierarchy concerning the value of 'universal indefiniteness' and its relation to any other indefiniteness value with regard to borrowing.

We first discuss borrowed indefiniteness **markers**. Internal grammaticalisation of borrowed elements (including borrowed determiners) into indefi-

niteness markers is not considered here (see Section 19.4). We also do not include indefiniteness markers that occur only in loan word-forms, and are not extended to indigenous bases. Finally, we consider only instances that retain the indefiniteness value of the marker as it occurs in the source language. For example, the widespread specific prefix *vare-* is not considered to be a loan of a specific marker, since it was not borrowed as such: it was borrowed as a free-choice marker and has acquired its specific function through internal developments (see Section 19.2). Some of the markers given in Tables 19.3–19.5 are applied to a whole series of indefinites, or at least to a few ontological categories. Others are restricted to a single ontological category, mostly the determiner (see Chapter 20).

Although not all dialects possess a distinct free-choice series of indefinites and although indefinites of this function are the worst attested in our data, borrowed free-choice markers are still very frequent and diverse. They are charted in Table 19.1 according to their source language. Only basic forms of the markers are given (e.g. *-godi* also represents *-gudi*, *-gode*, *-god* etc.). Some of the free-choice markers cover a wide range of further functions (e.g. German *irgend-*, Haspelmath 1997b: 245).

The free-choice markers usually originate in the current or a relatively recent L2 of the dialect, with the following exceptions: the South Slavic suffix *-godi*

Table 19.1. Borrowed markers of free-choice indefiniteness

Source L2	Marker	Dialects
South Slavic	*-godi*	Kosovo Bugurdži, Xoraxane, Slovene R, North Vlax
Serbian/Croatian	*bilo-*	Kosovo Bugurdži
	ma-	Gurbet
	makar-	Bosnian Gurbet, Central Slovak Romani
Rumanian	*fije-*	Bosnian and Kosovo Gurbet, Dasikano, Xoraxane, Rakarengo
	ori-	Rakarengo
	vare-	Welsh R, older Finnish R
Hungarian	*akār-*	South Central, Hungarian Vlax, Hungarian Sinti
Slovak	*bārs-*[a]	West Slovak R
Slovak; Polish	*xoč-*	East Slovak R; South Polish R; Latvian (Curland) R
German	*irgend-*	Austrian Sinti
Finnish	*vaxxa-*	Finnish R

[a] In Slovak (as well as in Czech, Slovene, and Croatian) dialects, this prefix is borrowed from Hungarian.

Table 19.2. Borrowed markers of negative indefiniteness

Source L2	Marker	Dialects
South Slavic	*ni-*	numerous (see below)
North Slavic	*ňi-* (*n'i-*)	Slovak/Czech Central; Northeastern, Crimean R
Turkish	*hidž-*[a]	Florina Arli, Sepečides
Azeri	(*heš-*)	Iranian R
Albanian	*as-*	Kosovo Gurbet

[a] In Turkish and Azeri, this prefix is borrowed from Persian.

in North Vlax, the Serbian/Croatian prefix *makar-* in Central Slovak Romani, and the Rumanian prefixes *fije-* and *vare-* in all dialects indicated. The Rumanian prefix (*v*)*oare-* is the source of *vare-* in numerous dialects in and outside of Rumania. However, it is rarely attested in its original free-choice function (see Section 19.2).

Borrowed negative markers are charted in Table 19.2. Again, only basic forms of the markers are given.

The South Slavic prefix *ni-* shows the widest cross-dialectal distribution. It is found not only in dialects in current contact with a South Slavic language – Prekmurje Romani, Slovene Romani, numerous Balkan dialects (Arli of Prizren, Gilan and Prilep, Sofia Erli, Yerli, Varna and Kosovo Bugurdži, Malokonare, Muzikanta, Nange, Drindari, Montana Kalajdži), Gurbet-like Vlax dialects (e.g. Gurbet, Dasikano, Xoraxane), and Serbian Kalderaš – but also in dialects that lost the contact with South Slavic a long time ago. These latter include Abruzzian Romani, most South Central dialects (those in current contact with Hungarian and German), Gurvari, and originally also Iranian Romani.[9] Dialects that, in all likelihood, also once possessed this prefix (South Central in current contact with Slovak, and Crimean Romani) readily adopt the palatal or palatalised North Slavic form. The Turkic and Albanian markers originate in current or recent L2s.

Borrowed specific markers are shown in Table 19.3. Only basic forms of the markers are given. Specific markers that arose through contamination of an older marker by the marker of the current L2 (e.g. *dare-* and *vale-* in some Central dialects developed from the older *vare-* influenced by Slovak *da-* and Hungarian *vala-*, respectively) are not shown. Borrowed specific markers usually originate in current or recent L2s.

Further, Bosnian Gurbet and Xoraxane have borrowed the negative polarity prefix *i-* from Serbian/Croatian; and Lithuanian, Russian, Ukrainian, and Crimean Romani have borrowed the negative polarity suffix *-n'ebud'* from

Table 19.3. Borrowed markers of specific indefiniteness

Source L2	Marker	Dialects
Greek	*kan-*[a]	Florina Arli, Rumelian R
Bulgarian	*-si*	Vălči Dol
Serbian/Croatian	*ne-*	Kumanovo Arli, Gurbet, Serbian Kalderaš
Rumanian	*-va*	Rakarengo
Albanian	*di-*	Arli of Gilan and Prizren, (Slovene R)[b]
	-far	Kosovo Bugurdži
Hungarian	*vala-*	South Central
Polish	*-ś*	some Polish R
Ukrainian	*-s'*	Ukrainian Romani
Ukrainian	*d'e-*	Ukrainian R
Russian	*-to*	Estonian, Lithuanian, Crimean R

[a] The prefix *kan-* occurs in the Romani determiner/person indefinite *kan-ek* 'some; somebody', possibly a calque on Greek *kan-enas*. However, since the Greek indefinite covers irrealis-to-negation functions, and not specific indefiniteness (Haspelmath 1997b: 265–266), the proposed etymology is doubtful. For an alternative etymology see Chapter 20.
[b] The Albanian origin of the prefix *di-* in Slovene Romani is doubtful.

East Slavic. Borrowing of universal markers appears to be rare: the only attested instances are *minden-* from Hungarian in Šóka Rumungro (e.g. *minden-kāj* 'everywhere') and Gurvari (e.g. *minden-ko* 'everybody'), and *har-* from Persian in Iranian Romani and *her-*[10] from Turkish in Ajia Varvara (e.g. *her-kon* 'everybody').

The asymmetry with regard to borrowing of indefiniteness markers may be partly stated in implicational terms: specific → negative (→ free-choice). Although universal markers are rare, they may be the only indefiniteness markers to be borrowed (e.g. in Ajia Varvara). Four types of dialects are shown in Table 19.4.

Table 19.4. Patterns of borrowing of indefiniteness markers

	Free-choice	Negative	Specific
Type A: Gadžikano	–	–	–
Type B: Welsh R	*vare-*	–	–
Type C: Dasikano	*fi-*	*ni-*	–
Type D: Klenovec Rumungro	*akār-*	*ňi-*	*vala-*

Dialects of Type A do not possess any borrowed indefiniteness marker. Type B (only free-choice markers borrowed) is, apart from Welsh Romani, attested in Hungarian and Austrian Sinti, Lovari and Taikon Kalderaš. Type C (free-choice and negative markers borrowed) is, apart from Dasikano, found in Latvian Romani, the North Central dialects, Kosovo Bugurdži, Serbian Kalderaš, and possibly also in Slovene Romani. Type D (free-choice, negative, and specific markers borrowed) is attested in the South Central dialects. Xoraxane, Bosnian Gurbet, Lithuanian, Russian, Ukrainian, and Crimean Romani roughly fit this type as well, except that they borrow a negative polarity marker instead of a specific one. The reason we put free-choice into brackets in the above implicational asymmetry is that numerous dialects do not possess a distinct free-choice series. In these dialects, only a negative marker is borrowed, or both a negative and a specific one (i.e. specific → negative).

There appear to be few absolute constraints on borrowing of indefinite **word-forms**. Nevertheless, free-choice indefinites are rarely attested (e.g. *štogod* 'whatever' from Serbian in Kosovo Bugurdži, and *akārmikor* 'whenever' from Hungarian in Šóka Rumungro). The rarity of the free-choice loans may be partly due to gaps in our data. We first discuss indefinite loans by ontological values, and then formulate some generalisations.

The most frequently borrowed distributive determiner is the Balkan Slavic universal *vsjako* (*vseko*, *svako*, *sjako*, *sako*, *seko*) 'every', which is found in the majority of Romani dialects. Further universal determiners are: *her* (*xer*, *er*) from Turkish in numerous dialects of Bulgaria and Turkey; *každo* or *kažno* from North Slavic in Polish, Lithuanian, and Crimean Romani; and *kathe* from Greek in Karditsa.[11] Borrowed specific determiners include: *njakakvo* (*nekakvo*) 'some' from Bulgarian in Montana Kalajdži and Gadžikano; *neko* from Serbian in Kosovo Bugurdži; *filjan* (*filani*) from Turkish in Ajia Varvara, and from Albanian in Kosovo Bugurdži; *jenego* from German in Manuš; and *joku* from Finnish in Finnish Romani. There are also borrowed specific determiners used with plural head nouns: *bazi* from Turkish in Kaspičan and Karditsa; *kapja* from Greek in Karditsa; and *uni* (*vuni*) and *nešte* (*nište*, *mište*) from Rumanian in the North Vlax dialects. A loan of a specific determiner seems to imply a loan of a universal determiner. Borrowed negative determiners are extremely rare: the only attested example is *žiadno* 'no, none' from Slovak in some Central dialects.

Borrowed person indefinites are relatively rare; some of them also function as determiners. Greek is the source of the specific *kapjos* 'somebody' in Karditsa and Ajia Varvara, and the universal *kathenas* in Ajia Varvara. Macedonian or Bulgarian provided the forms *njakoj* (*nekoj*) 'somebody', *nikoj* 'nobody'

and *sjakoj* (*sekoj*, *vsekoj*) ‘everybody’. All three indefinites are attested in Prilep Arli and Varna Bugurdži, while Sofia Erli and Yerli lack the specific loan, and Kumanovo Arli lacks the universal loan. The negative *niko* ‘nobody’ in Prizren Arli and the negative-polarity *iko* ‘anybody’ in Xoraxane may be internal de-interrogative derivations by means of borrowed indefiniteness markers (see above), as well as loans from Serbian/Croatian. Most dialects outside of the Balkans have borrowed the universal *sako* (*svako*) ‘everybody’ from Serbian/Croatian, and in some of them (e.g. the Central and the Northeastern dialects) it is the only borrowed person indefinite. Finnish Romani of Kuopio has *joku* ‘somebody’ from Finnish. There seems to be no generally valid borrowing asymmetry for person indefinites.

Slavic loans are the most frequent among borrowed thing indefinites: cf. the negative *ništa* (*ništo*, *nič*, *ňišt*, *ňič*) ‘nothing’ from South Slavic or Slovak in the Central dialects, Slovene and Abruzzian Romani, some Balkan dialects (e.g. Arli, Sofia Erli, Yerli, Varna Bugurdži, Crimean Romani, Kosovo Bugurdži, Nange), Xoraxane and Serbian Kalderaš; the specific *nešto* from Macedonian or Bulgarian in some Gurbet varieties, Arli, and Varna Bugurdži; and the universal *svašta* ‘everything’ in some Gurbet varieties and Serbian Kalderaš.[12] Greek has provided the specific *kati* ‘something’ in Karditsa, Sepečides and Ajia Varvara and the irrealis-to-negative *tipota* ‘something, anything, nothing’ in Karditsa and originally also in Iranian Romani.[13] Further source languages include Hungarian (*minden* ‘everything’ in Gurvari), German (*lautə* ‘everything’ in some Sinti varieties), and Finnish (*jotain* ‘something’ in some modern varieties of Finnish Romani). Negative loans are more frequent than specific loans, and the latter tend to imply the former. Universal loans are relatively rare, but they might be the only thing indefinites to be borrowed (e.g. in Sinti). Borrowed thing indefinites frequently originate in the current L2, but there are numerous exceptions: especially the South Slavic *ništa* ‘nothing’ tends to be retained for a long time (e.g. in the South Central dialects, Abruzzian Romani, or Crimean Romani).

Borrowed place indefinites, which are common only in the Balkans, always originate in the current L2. South Slavic loans are by far the most frequent: some dialects (e.g. Slovene Romani, Prizren Arli, Sofia Erli, Yerli, Kosovo Bugurdži, Montana Kalajdži, Dasikano, and Serbian Kalderaš) possess only the negative *nikəde* (*nigde*, *nigdi*, *nindźe*) ‘nowhere’, while others (e.g. some Arli varieties, Nange, Muzikanta, and Varna Bugurdži) have also borrowed the specific *njakəde* (*negde*) ‘somewhere’. The universal *svugde* (*segdeka*) is attested in some Gurbet varieties and in Prilep Arli. Other source languages are Greek (cf. the specific *kapu* ‘somewhere’ in Karditsa and Ajia

Varvara) and German (cf. the universal *ivral* 'everywhere' in some Sinti varieties).[14] The Slavic loans seem to indicate that borrowing of specific place indefinites implies borrowing of their negative counterparts. However, the Greek specific loan without a negative counterpart contradicts this generalisation.[15] The Sinti case shows that a universal place indefinite may be the only one to be borrowed (cf. the indigenous *kaj-komuni* 'somewhere' and *kajni* 'nowhere').

Borrowing of time indefinites is perhaps the norm in Romani. They usually originate in the current L2 and, more rarely, in a recent L2. Borrowed negative indefinites ('never') include: *pote* from Greek in Ajia Varvara; *nikoga*(*š*) from Macedonian or Bulgarian, *nikad* from Serbian/Croatian, or *nindar* from Slovene in most dialects in contact with these languages; *kur* from Albanian in Priština Gurbet; *asla* from Turkish in Razgrad Drindari; *šoha* from Hungarian in many Central dialects, Lovari, and Taikon Kalderaš; *n'ik'edi* (*n'igdi*, *n'igda*) or *n'ikol'i* from North Slavic in many North Central and most Northeastern dialects; *ni*(*a*) from German in many Sinti varieties; *žame* from French in Manuš; and *maj* or *džamaj* from Italian in Piedmontese Sinti and Abruzzian Romani. Borrowed universal indefinites ('always') include: *vinagi* from Bulgarian, *sekoga*(*š*) from Macedonian, *uvek* from Serbian/Croatian, or *vavik* Slovene in most dialects in contact with these languages; (*h*)*ep* from Turkish in Florina Arli, Sepečides, Kaspičan, and Gadžikano; *mindig* from Hungarian in some Central dialects and Lovari; *žd*i from Slovak in some Central dialects; *furt* from Hungarian or Slovak[16] in the Central dialects; *zavše* from Polish in Polish Romani; *vs'egda* (*sagda*) from Russian in Lithuanian Romani and Taikon Kalderaš; *imer* from German in some Sinti varieties; and *alti* from Swedish in Finnish Romani. Borrowed specific indefinites ('sometimes') are rarer: cf. *njakoga*(*š*) from Macedonian or Bulgarian, *nekad* from Serbian/Croatian, or *učasi* from Slovene in Macedonian Gurbet, Arli of Kumanovo and Prilep, Varna Bugurdži, Malokonare, and Slovene Romani; *dikur* from Albanian in Kosovo Bugurdži; *valamikor* from Hungarian in some South Central dialects; and *manxmal* from German in some Sinti dialects. Borrowed specific indefinites imply loans of their negative counterparts. There seems to be no implicational asymmetry between negative and universal loans, which are equally frequent.

Borrowing of manner indefinites (e.g. Prilep Arli *nikako* 'nohow', *nekako* 'somehow', and *sekako* 'in all ways' from Macedonian) is rare or rarely attested, and so we cannot formulate valid generalisations. The same holds for quantity indefinites (e.g. Crimean *n'eskol'ko* 'some [amount of]' from Russian, and Montana Kalajdži *njakolko* from Bulgarian).

The discussion of borrowing of indefinite word-forms may be summarised as follows. First, negative indefinites are, on the whole, more likely to be borrowed than specific indefinites. This may be formulated as an implication (specific → negative), at least with thing and time indefinites. A significant exception to this generalisation is observed in indefinite determiners where the opposite asymmetry holds: specific determiners are more likely to be borrowed than negative determiners.[17] Second, borrowed universal indefinites are very frequent: roughly as frequent as borrowed negative indefinites, and more frequent than borrowed specific indefinites. However, only with determiners and time indefinites, a loan of a specific indefinite implies a loan of a universal indefinite.

Chapter 20
Ontological category

By 'Ontological' we mean the semantic domain specification that is assigned to grammatical operations. The ontology of a grammatical expressions is the domain in which the operation that is triggered by that expression is valid. Thus, an interrogative conveys a general instruction, in communicative terms, to add or supplement information. The ontological specification narrows down the semantic-conceptual or real-life domains to which this information may belong: persons, things, places, and so on. Our aim in this chapter is to investigate whether different ontological domains are prioritised in different ways. The ontological category is encoded in pro-words, especially in interrogatives and indefinites, but also in deictics[1] and some other pro-words. The ontological values include: determiner, person, thing, place, time, manner, cause, goal, quantity (amount), and size. They are illustrated by "typical" Romani interrogatives and their English translations in Table 20.1.

Size pro-words are only attested in a few dialects (e.g. Šóka Rumungro *kibedor* 'how big', *ebedor* 'that big'); they are mostly replaced by periphrastic constructions such as *savo baro* [which/what.sort.of big] 'how big'. Most dialects do not distinguish cause and goal, encoding both by *soske* 'why' or its derivations (e.g. *varesoske* 'for some reason, on some grounds'). If the

Table 20.1. Ontological values in Romani interrogatives

Value	Romani	English
determiner	*savo*	'which, what sort of'
person	*kon*	'who'
thing	*so*	'what'
place	*kaj*	'where'
time	*kana*	'when'
manner	*sar*	'how'
cause	(*sostar*)	'why, on what grounds'
goal	*soske*	'why, what for'
quantity	*keti*	'how many/much'
size	(*kibor*)	'how big'

two functions are distinguished (e.g. in Florina Arli, North Vlax, and partly in Kaspičan), then the forms based on *soske* encode goal, while forms based on *sostar* encode cause. Some dialects distinguish between quality determiners and identification determiners (e.g. Kalderaš *savo* 'which' vs. *če* or *sosko* 'what sort of'). However, most dialects do not encode this distinction (e.g. East Slovak Romani *varesavo* 'some sort of, some [individual/s of]'). We will mostly discuss asymmetries between the major ontological values (viz. determiner, person, thing, place, time, and manner), only sometimes taking into account cause, goal, and quantity, and rarely size.

There are multiple cross-cutting categories for the ontological category, including: lexical type of the pro-word (interrogative, indefinite, deictic, other); indefiniteness, with indefinite pro-words; deictic distinctions, with deictics; orientation, with place pro-words; and, again, lexical type, with quantity pro-words (cardinal, ordinal, multiplicative). For example, there are at least 36 place pro-words in Šóka Rumungro, including *akārkathar* 'from anywhere whatsoever' (a free-choice indefinite of separative orientation), *okija* 'just to there' (a specific remote deictic of directive orientation), or *āvermerre* 'though somewhere else' (an *else*-pro-word of perlative orientation). Due to gaps in our data as well as practical limitations we will mainly discuss ontological asymmetries in interrogatives and indefinites.

The category of animacy in nouns is closely related to the ontological category, and so we include it in this chapter. While nouns denoting humans may be assigned the person value and nouns denoting inanimate objects may be assigned the thing value, non-human animates (especially animals) fall in between. Nevertheless, it turns out that this value is not required by our data (see Section 20.1), and so we leave it out of consideration.

The determiner value is likely to be borrowed, relatively likely to extend, showing medium differentiation and internal diversity, and not likely to erode or exhibit extracategorial distribution. There is conflicting evidence with regard to its complexity. The person value is the most inflectionally differentiated, relatively likely to extend, of medium internal diversity, not very complex, and not likely to be borrowed or show extracategorial distribution. The thing value has the widest extracategorial distribution, is highly differentiated, may extend, is not very complex, not very likely to be borrowed, and not very likely to be eroded, and it is not at all internally diverse. Animates (corresponding to the person value) are more complex than inanimates (corresponding to the thing value). The place value is very diverse and likely to show extracategorial distribution, is commonly borrowed, may extend, and is not very complex. The criterion of differentiation gives conflicting results: the place

value is highly differentiated in orientation, but has no inflectional differentiation. The time value is the most likely to be borrowed, shows medium extracategorial distribution, does not extend, and is not differentiated or internally diverse. There is conflicting evidence with regard to its complexity. The manner value shows medium complexity, medium tendency to erode, and may be distributed extracategorially; it is rarely borrowed, not at all internally diverse or differentiated, and does not extend. The cause/goal value is the most likely to erode, very complex, not very likely to be borrowed, shows little internal diversity or extracategorial distribution, does not extend, and is not differentiated. Finally, the quantity value is the most internally diverse, shows some differentiation and external distribution, is not very complex or likely to be borrowed, and does not extend.

It is clear from the above overview that there is no single ontological hierarchy. In fact, if we just consider those criteria that clearly assign the greatest prominence to a single value, we find that four different values are selected by four criteria: thing by extracategorial distribution, time by borrowing, cause/goal by erosion, and quantity by internal diversity.

20.1. Complexity

Most indigenous interrogatives are bimorphemic (in their base forms), containing an interrogative root *k-* or *s-* and a suffix that encodes the ontological value. The determiner *s-av-o* is trimorphemic. The (trimorphemic) cause/goal interrogatives *so-s-ke* 'why; for what' and/or *so-s-tar* 'on what grounds' are dative or ablative forms of the thing interrogative *so* 'what', and may be considered to be synchronically derived from it.[2] This is also the case with the (quadrimorphemic) North Vlax determiner *so-s-k-o* 'which', which is the genitive form of the thing interrogative. It is likely that the indigenous manner and determiner interrogatives are also historical derivations of the thing interrogative (cf. *s-o* > *s-ar* 'how' and *s-av-o* 'which, what sort of'). The complexity hierarchy with indigenous interrogatives is thus: determiner (tri- or quadrimorphemic, derived) > cause/goal (trimorphemic, derived) > manner (bimorphemic, derived) > person, thing, time, quantity (bimorphemic). The position of the place value on the hierarchy is ambiguous due to differing complexity of interrogatives of different orientation (see Chapter 18 for details).

The hierarchy found in interrogatives, including the greatest complexity of the determiner, also holds for de-interrogative indefinites. However, in many

Table 20.2. Determiner-base indefinites in selected dialects

	Sepečides	Muzikanta	Florina Arli	Kos. Bugurdži
Determiner	*hidžekh*	*kek*	*kane*	*haj*
Person	*hidžekh dženo*		*kane* (*dzeno*)	*haj dženo, haj-ek*
Thing	*hidžekh-šej*	[*čhipas*]	[*čumuni*]	*haj-či*
Place	*hidžekhe thaneste*	*keki thaneste*	*kane thane*	*haj-gode*
Time	*hidžekh-far*	*kek vakəci*	*kane fora*	[*dikur*]

dialects, especially those of the Balkans, the indefinite determiner tends to be the least complex value. Consider the selected indefinites in four dialects in Table 20.2 (negative in Sepečides, specific in the other dialects).

In most instances in Table 20.2, the indefinites consist of the determiner plus a base that indicates their ontological value; we will call them *determiner-base* indefinites. Here, the determiner functions as an indefiniteness marker, or better: it consists of an indefiniteness marker alone, without containing an overt base, and so it is the least complex value. Muzikanta and Florina Arli illustrate that the person indefinite may be identical to the determiner (see also Section 20.4), and so less complex than indefinites of the other ontological values. Hence the complexity asymmetry in these dialects: other > person > determiner. There does not seem to be any obvious asymmetry between the other ontological values (thing, place, and time).[3] The ontological base of determiner-base indefinites is a more or less grammaticalised generic noun such as 'person' (*dženo*), 'human' (*manuš*, *menšo*), or 'soul' (*zelo*) for person; 'thing' (*šej*, *idos*) for thing; 'place' (*than*, *jer*, *stedos*) for place; and 'time' (*vakəti*, *vreme*, *cajto*) or 'day' (*dives*, *gün*) for time.[4] In some instances, the base has fused with the determiner (e.g. Taikon Kalderaš *katende* or Malokonare *kacinende* 'somewhere' < **kaj thanende* 'in some places'). Generic bases for thing and time are restricted to the Balkans, while place and person generic nouns are also found in some Northwestern dialects. The Kosovo Bugurdži examples in Fig. 20.2 illustrate that the ontological base may also be the numeral 'one' (as in *haj-ek* 'somebody' < **kaj-jekh*), a former indefinite (as in *haj-či* 'something'), or a borrowed interrogative base (as in *haj-gode* 'somewhere'). Widespread is the use of multiplicative markers as temporal bases: the indigenous **-var* (e.g. Taikon Kalderaš *vuni-var* 'sometimes', Sepečides *hidžekh-far* 'never', or Welsh Romani *kekār* 'never' and Finnish Romani *čekkar* '(n)ever' < **kek-var*) and others (e.g. Florina Arli *bazi fora* 'sometimes', Muzikanta *xer drom* 'always', Polish Romani *każdo moło* 'always').

Table 20.2 also shows that various structural types of indefinites may combine within a dialect. For example, the thing indefinites in Muzikanta and Florina Arli are not of the determiner-base type, although the other indefinites are. This makes it difficult to evaluate complexity differences between ontological values within a dialect, and especially across dialects. To sum up, the hierarchy in (1) holds for interrogatives and de-interrogative indefinites, while the hierarchy in (2) holds for determiner-base indefinites. Note especially the conflicting position of the determiner.

(1) Determiner > cause/goal > manner > person, thing, time, quantity
(2) Other > person > determiner

As for animacy, animate nouns tend to be more complex in respect of case marking. Most dialects show a split between the marking of the direct object, with animates taking the accusative (i.e. the independent or markerless oblique), and inanimates taking the nominative. In case relations that are represented by adpositions, the adposition may in some dialects govern an adpositional case (usually locative) with animate nouns, and nominative with inanimates. Holzinger (1993) argues against a polarised animacy scale with just two values at its extreme ends – animate and inanimate, and proposes instead an animacy continuum, with humans ranking highest, and various classes of animals occupying intermediary positions. While we would not dispute such a hierarchy, it is not apparent in our sample, however, and most dialects show the following, polarised pattern (3):

(3) Slovak Romani (Lučivná)
   a. *Dikhľas ole phure muršes.*
      saw.3SG that.OBL old.OBL man.ACC
      'He saw the old man.'
   b. *Leskro dad murdardžas kole grajes.*
      his father killed.3SG that.OBL horse.ACC
      'His father killed that horse.'
   c. *Murdardžas la kaxňa la čhuraha.*
      killed.3SG the.OBL chicken.ACC the.OBL knife.INSTR
      'He killed the chicken with the knife.'
   d. *Dikhľom oda kher.*
      saw.1SG that.NOM house.NOM
      'I saw that house.'

An exception to the pattern is the Vălči Dol dialect, which appears to have generalised nominative marking with all nominal direct objects. Despite the general animacy split in direct object marking, however, numerous instances of animate direct objects in the nominative can be found, in the elicited corpus and in the literature, from almost all dialect groups. Factors that seem to promote the choice of the nominative are indefiniteness, and in particular the introduction of non-specific, unidentified animate entities (4)–(5):

(4) Austrian Lovari (Cech and Heinschink 1998: 36)
Lel peske gažo.
take.3SG REFL.DAT husband(.NOM)
'S/he is getting married.'

(5) Serbian Kalderaš (Boretzky 1994: 101)
Avili i vrjamja laki śej te lel gaźo.
came.3SG ART time her daughter COMP take.3SG husband(.NOM)
'The time came for her daughter to get married.'

Another promoting factor is the appearance of the animate direct object in an ambiguous position, as a possible subject-topic of the following clause or predication (6)–(8):

(6) Austrian Sinti
Auf amol dikeles lako tikno phral buter nit koj.
suddenly see.3SG.REM her little brother more not there
'Suddenly she sees her little brother no longer there.' (viz. 'she no longer sees her little brother there', or 'she sees her little brother is no longer there')

(7) Varna Bugurdži
Dikljom manuša tjorna anglal ki magazina.
saw.1SG humans stand.3PL in.front.of at shop
'I saw men standing in front of the shop.'

(8) Sofia Erli
Me dikhljom ki ulica jekh mruš te phirel.
I saw.1SG at street one man COMP walk.3SG
'I saw a man walking down the street.'

Table 20.3. Marking of nouns in adpositional case role ('behind')

	Type A	Type B	Type C	Type D	Type E	Type F
Animate	LOC	ACC	*ko* + NOM	NOM	**GEN + POST**	LOC (ABL)
Inanimate						NOM

Contrasting with the extension of the nominative to animate direct objects, there are no examples of oblique marking of inanimate direct objects (but see Boretzky 1994: 102 for an exception). In other thematic roles non-nominative marking of inanimate objects is more widespread. Table 20.3 summarises the principal options for case marking, here with the adposition *pal* 'behind'.

The conservative patterns (types A and B) are found predominantly in the Balkans. They show, irrespective of animacy, non-nominative marking of the prepositional object. In Type A (e.g. Varna Bugurdži, Varna Kalajdži, Sliven, Kaspičan, Malokonare, Yerli), the object is in the locative (9), while in the less frequent Type B, it is in the accusative (10):

(9) Varna Kalajdži
a. *O cəkno xurdo garadilo pala eke kopačeste.*
the small.M boy hid.3SG.M behind one.OBL tree.LOC
'The little boy hid behind a tree.'
b. *Voj pirelas pal eke muršəste.*
she walk.3SG.REM behind one.OBL man.LOC
'She was walking behind a man.'

(10) Šumen Drindari
a. *O cikoro kəzəes garaes pe pala i kaštes.*
ART small.M boy hid.3SG REFL.ACC behind the.OBL tree.ACC
'The small boy hid behind the tree.'
b. *Oj phirlas pala ek romes.*
she walk.3SG.REM behind one man.ACC
'She was walking behind a man.'

Type C (Karditsa Arli and Crimean Romani) shows another preposition, *ko*, derived from an element that mirrors a Layer II marker, which mediates between the location-specific preposition *pala*(*l*) and the noun (e.g. Crimean Romani *pal ke murš* 'behind the man'). This too is in all likelihood an archaic feature, though in Type C the construction governs the nominative and not the

accusative, which was possibly the original Early Romani case in this construction of Adverb+Preposition+Noun+Layer I case (cf. similar layout, albeit with postposed free markers, in subcontinental New Indo-Aryan). Type D is widespread among the Vlax, Balkan, and Central dialects (e.g. Sofia Erli, Lovari, Kalderaš, Slovak Romani, Klenovec Rumungro, Rešitare, Vălči Dol, and more), showing complete reduction of nominal case with the preposition. Type E, showing genitive marking of the head and a postposition, is restricted to Finnish Romani (*jeko jēnesko pālal* [one person.GEN behind] 'behind a man'). Some Rumungro varieties show a similar construction, albeit with a preposition, genitive marking of the head, and the noun 'back' as a spatial metaphor following the head. Animacy split is thus restricted to Type F, comprising just the Northeastern dialects (Polish Romani showing replacement of the locative by the ablative case).

20.2. Erosion

There appears to be a single salient erosion development in Romani which is relevant for ontological asymmetries. In some dialects, the initial interrogative root *s-* (as in *s-o* 'what', *s-avo* 'which', *s-ar* 'how' and *s-oske* 'why') has been eroded to *h-* (or later to zero). Table 20.4 shows the distribution of the interrogative root in five types of dialects.

While most Romani dialects possess only the root *s-*, and some Core Sinti dialects possess only the root *h-*, there are three types of dialects with some interrogatives in *s-* and some in *h-*. The root *h-* is most common in the cause/goal interrogative, less so in the manner interrogative, and the least common in the determiner and the thing interrogatives. Two unilateral implications concerning obligatory presence of *h-* may be formulated: (a) thing → manner → cause/goal and (b) determiner → cause/goal. The implication (c) determiner

Table 20.4. Patterns of erosion of the interrogative root s-

	Thing	Determiner	Manner	Causal
most dialects	*s-*	*s-*	*s-*	*s-*
Piedmontese Sinti	*s-*	irrelevant	*s-*	*s-* ~ *h-*
western North Central	*s-*	*h-*	*s-* ~ *h-*	*h-*
Sinti (Austria, Germany)	*h-*	*s-*	*h-*	*h-*
Sinti (Hungary, Manuš)	*h-*	*h-*	*h-*	*h-*

→ manner concerns optional presence of *h-*. The implied ontological values are more likely to undergo erosion of the root. There are conflicting asymmetries between the thing and the determiner interrogatives (e.g. Austrian Sinti *ho* 'what' and *saw* 'which', but Bohemian Romani *so* and *havo*).

20.3. Differentiation

Place pro-words are highly differentiated by the cross-cutting category of orientation, and quantity pro-words show differentiation into cardinals, ordinals, and multiplicatives (e.g. Slovak Romani *ajci* 'that much', *ajci-to* 'in that place in an order', and *ajci-var* 'that many times'). Below we discuss ontological asymmetries concerning declinability, inflectional differentiation, and inflectional irregularity.

Pro-words of person and thing show substantival inflection. Adjectival pro-words of different ontological value differ in the extent of inflection. Pro-words of ordinal quantity always inflect as adjectivals. Most determiners (e.g. *savo* 'which', its indefinite derivative, *sako* 'every', demonstratives) show adjectival inflection, while other determiners (e.g. *kaj* 'some', *če* 'which') are indeclinable modifiers. Pro-words of cardinal quantity are always indeclinable as modifiers, and they only inflect when substantivised (e.g. *keti džen-enca* 'with how many people' vs. *ket-enca* 'with how many of them'). Adverbial pro-words, viz. those of place, time, manner, cause/goal, and multiplicative quantity, are always uninflected.

Substantival inflection is much more differentiated than adjectival inflection in case (see Chapter 21), and so person and thing pro-words are more differentiated than the other inflected pro-words. Moreover, thing pro-words (like other nominals referring to inanimates) lack a distinct accusative case, and so person pro-words are more differentiated. On the other hand, inflected adjectival pro-words encode number and gender, which is only rarely encoded in person pro-words and almost never in thing pro-words. Since substantival case establishes more distinctions than number and gender in adjectival paradigms, substantival pro-words are on the whole more differentiated than adjectival ones. The criteria of declinability and inflectional differentiation thus render the following ontological asymmetry in pro-words: person > thing > ordinal quantity > determiner > cardinal quantity > place, time, manner, cause/goal, multiplicative quantity.

Moreover, person pro-words show inflectional irregularities in some dialects. Thus, the person interrogative has irregular accusative and instrumental

forms in the South Central dialects (e.g. *kasaha* 'with whom' instead of the regular **kaha*), and the person indefinite *khonik* 'somebody; nobody' undergoes an irregular vowel alternation in its oblique stem in some Vlax dialects (e.g. Kalburdžu *konik* > *kanik-a-*). Irregularities in thing pro-words are rarer (e.g. the reduplication of the indefinite *či* 'something; nothing' in its oblique stem), and the inflection of adjectival pro-words is mostly regular.

20.4. Extension

Extensions most commonly concern the determiner and the person values; both directions of extension are attested: determiner to person as well as person to determiner. Extension of other types are rare: determiner to thing, person to place, thing to cause/goal, and place to determiner.

The universal determiner *sako* (*svako* etc.) 'every' is generally also used in head positions as a universal person indefinite 'everybody'. Only a few dialects distinguish between this determiner and the corresponding person indefinite (e.g. Austrian Sinti *sako* 'every' vs. *sakano* 'everybody'). Similarly, dialects that possess the universal determiners *savoro* (*saro*, *havoro* etc.) 'all, whole' and/or *celo* (*calo*, *cilo*) 'whole' frequently extend their plural forms to the person value to mean 'everybody'. The determiner *savoro* may also extend to the thing value to mean 'everything'.

As for specific and/or negative indefinites, numerous dialects show a formal affinity between the determiner and the person values. One may distinguish between two structural types of forms: simple forms (e.g. *kaj*, *khaj*, or *haj* in Kosovo Bugurdži, Kalderaš or Ajia Varvara, or *daj* in Prizren Arli), and forms based on the numeral **jekh* 'one'. The latter might simply correspond to the numeral (e.g. *jek* in some Sinti varieties, *ek* in Kaspičan and Gadžikano), or derive from it by means of an indefiniteness prefix. The indefiniteness prefix is either borrowed (e.g. *ni-jek* in Prizren Arli or *hidž-ek* in Sepečides, see Chapter 19) or, more commonly, grammaticalised from the simple indefinites: cf. **kaj-jekh* (e.g. *kajek* in Muzikanta, *hajek* in Kosovo Bugurdži, *k*(*aj*)*ek* in Malokonare, and Kalburdžu, *k*(*h*)*ak* in Kalderaš, *kek* in Welsh Romani and most Sinti varieties, *ček* in Finnish Romani) and **daj-jekh* (e.g. *d*(*aj*)*ek* in Kosovo Bugurdži and Prizren Arli, *dekh* in Sofia Erli, and probably also *tek* in some Sinti varieties); there are also some less transparent forms of this type (e.g. *kan-ek* in Florina Arli and Rumelian Romani, and *b-ek* in Yerli).[5] Now, while the simple forms always function as determiners, a *jekh*-based form may be either only a person indefinite (e.g. Kosovo Bugurdži *haj* 'some' vs. *hajek*

'somebody') or both a determiner and a person indefinite (e.g. Welsh Romani *kek* 'no, none; nobody'). If a *jekh*-based form is only used as a determiner, then the corresponding person indefinite consists of this determiner plus a generic noun (e.g. Malokonare *kek* 'some' vs. *kek manuš* 'somone') or derives from the determiner (e.g. Sofia Erli *dekh* 'some' vs. *dekh-oj* 'someone'). The distribution of the *jekh*-based forms appears to imply that, diachronically, they extend from the person value to the determiner value.

The person indefinite *komonī* 'someone' is also used as a place indefinite 'somewhere' in Welsh Romani. In Austrian Sinti, *komuni* is used as a base for the place indefinite *kaj-komuni* (literally 'where-someone'). Numerous dialects employ the thing interrogative *so* alongside the specifically cause/goal interrogative *soske* in a causal sense (e.g. *so asas?* 'why are you laughing?'), due to convergence with various contact languages. The place interrogative *kaj* is used as a determiner in a few dialects of the Balkans (see also Section 20.6).

20.5. Extracategorial distribution

Interrogatives, which primarily occur in independent interrogative clauses and in interrogative embeddings, show extended distribution in that they are used as connectors in various subordinate constructions. In this section, we explore distribution asymmetries between interrogatives of different ontological values.

First, some interrogatives are used as **relativisers**.[6] The place interrogative *kaj* appears to be most widespread, and may be reconstructed for Early Romani, where it developed due to structural convergence with Greek. As a general relativiser it is attested in older Sinti, older Central dialects, the Northeastern dialects, Slovene Romani, some Balkan dialects (e.g. Arli of Prizren, Prilep and Florina, Sepečides, Sofia Erli, Crimean Romani, Varna Bugurdži, Nange, Gadžikano, and Kaspičan) and many South Vlax dialects (e.g. Xoraxane, Ajia Varvara, Varna Kalajdži, Rešitare, and Vălči Dol).[7] In a few dialects (e.g. Nange and Gadžikano), *kaj* now functions only as a relativiser, having been completely replaced in its interrogative function (see Chapter 18). Due to convergence with later contact languages, some dialects now also use the indigenous determiner, person, and thing interrogatives as relativisers. Depending on the dialect, the person interrogative is restricted to clauses modifying human nouns and the thing interrogative is restricted to clauses modifying non-human nouns, or there are no such categorical restrictions. Relativisers

of all these three ontological values are found in Finnish Romani, Piedmontese Sinti, South Central dialects, Slovene Romani, and Prilep Arli. In further dialects, only some of these relativisers are attested: the determiner and thing interrogatives in Slovak Romani; the determiner and person interrogatives in Crimean Romani, Xoraxane, and Vălči Dol; the determiner interrogative alone in the Northeastern dialects and Kalburdžu; and the thing interrogative alone in Austrian Sinti, Prizren Arli, Sofia Erli, and Austrian Lovari.

Second, some interrogatives are used as factual **complementisers**. The place interrogative *kaj* is most widespread in this function, again due to the structural convergence with Greek in the Early Romani period. Polish Romani, Bohemian Romani, Roman, and Florina Arli use *kaj* as a complementiser with epistemic verbs as well as some complex complementisers in manipulation complements and purpose clauses. In German Sinti, some Balkan dialects (e.g. Sofia Erli, Sepečides, Crimean Romani, and Kosovo Bugurdži) and some South Vlax dialects (e.g. Priština Gurbet, Ajia Varvara, and Varna Kalajdži), the place interrogative functions only as an epistemic complementiser. In some dialects of Slovakia (e.g. West Slovak and Lučivná Slovak Romani, and Klenovec Rumungro), on the other hand, the place interrogative is only found within complex manipulation and purpose complementisers. The Northeastern dialects (with the exception of Polish Romani) and Ukrainian Romani employ the thing interrogative *so* as a complementiser with epistemic verbs as well as within complex manipulation and purpose complementisers, due to convergence with East Slavic.

Third, interrogatives are used as subordinators in various adverbial clauses.[8] As for **temporal** clauses, subordination through indigenous or borrowed time interrogatives is the norm, and may be reconstructed for Early Romani. Only a few dialects (e.g. Austrian Sinti and the Northeastern dialects) have completely replaced the time interrogative by the manner interrogative in this function. The manner interrogative is also very common, being frequently employed to encode specific types of temporal subordination (especially punctual). Apart from the dialects mentioned above, it is attested in the Central dialects and many dialects of the Balkans (e.g. Arli of Prizren and Prilep, Sepečides, Crimean Romani, Kosovo Bugurdži, Gadžikano, Priština Gurbet, Ajia Varvara, and Rešitare). The place and thing interrogatives are also well attested in temporal clauses: the former[9] in Finnish, West Slovak, and Slovene Romani, Ajia Varvara, and some Balkan dialects (e.g. Florina Arli, Velingrad Yerli, Varna Bugurdži, and Malokonare), the latter in Polish and Slovak Romani, Sepečides and Kosovo Bugurdži. The quantity interrogative as a temporal subordinator is only attested in Arli of Prilep and Florina.

Interrogatives as **conditional** and concessive conditional subordinators mostly arose through extension of temporal subordinators, and they are much rarer than the latter. The time interrogative as a conditional subordinator is attested in Austrian Sinti, some Central dialects, and Kosovo Bugurdži, and as a concessive conditional subordinator in Slovene Romani. Polish Romani also shows the manner and the thing interrogatives functioning as conditional subordinators. However, the use of quantity interrogatives as concessive conditional subordinators in some Balkan dialects (e.g. Florina Arli, Sofia Erli, or Nange) seems to be partly independent of their use as temporal subordinators.

Various interrogatives are also used as adverbial subordinators (or as a part thereof) in clauses encoding **causal** relations (cause, reason, explanation and, more rarely, result). The cause/goal interrogatives *soske* or *sostar* are most common in this function, being found in many dialects of the Balkans (e.g. Arli of Prizren and Prilep, Sofia Erli, Yerli, Varna Bugurdži, Kosovo Bugurdži, Malokonare, Muzikanta, Gadžikano, Xoraxane, Ajia Varvara, Varna Kalajdži, Rešitare, Vălči Dol, and Kalburdžu) and marginally also in Roman.[10] The use of the place interrogative *kaj* as a causal subordinator is likely to be in fact a functional extension of the factual complementiser (e.g. in Roman, Florina Arli, Sofia Erli, and Crimean Romani). However, in some dialects (e.g. Slovene Romani, Velingrad Yerli, Varna Bugurdži, and Rešitare), *kaj* functions as a causal subordinator but not as a complementiser any more.

Finally, certain interrogatives are used as connectors[11] in **equative** and **comparative** constructions. The manner interrogative *sar* is generally used as a connector in equative constructions, and in many dialects (e.g. the Northwestern dialects, Estonian Romani, the Central dialects, and Lovari) also in comparative constructions. In numerous dialects of the Balkans, equative constructions may also contain the quantity interrogative (e.g. Prilep Arli *keti*, Varna Bugurdži *kozom*, Gadžikano *kirom*, or Rešitare *kobor*).[12] The thing interrogative *so* as an equative connector is only attested in Varna Kalajdži (e.g. *baro so tute* [big what you.SG.LOC] 'as big as you').

The asymmetries in (11)–(16) summarise the differing tendency of interrogatives of different ontological values to be employed as connectors in various subordinate constructions (on the basis of cross-dialectal frequency of occurence):

(11) Relative: place > determiner > thing > person > other
(12) Complement: place > thing > other
(13) Temporal: time > manner > place, thing > quantity > other
(14) Conditional: time > quantity > manner, thing > other

Table 20.5. Interrogatives as connectors in subordinate constructions

	Thing	Place	Man.	Quant.	Time	Person	Det.	Caus.
Relative	+	+	−	−	−	+	+	−
Complement	+	+	−	−	−	−	−	−
Temporal	+	+	+	+	+	−	−	−
Conditional	+	−	+	+	+	−	−	−
Causal	−	+	−		−	−	−	+
Equative	+	−	+	+	−	−	−	−
Constructions	5	4	3	3	2	1	1	1

(15) Causal: cause/goal > place > other
(16) Equative: manner > quantity > thing > other

Table 20.5 summarises the mere attestation of the extended distribution of interrogatives of different ontological values, and the number of constructions in which they may occur.

It is obvious that the ontological asymmetries are construction dependent, i.e. that there is no uniform asymmetry for the extended distribution of different ontological values. Nevertheless, the following global hierarchy may be formulated on the basis of their construction versatility (as counted in Table 20.5):[13]

(17) Thing > place > manner, quantity > time > cause/goal, determiner, person

20.6. Internal diversity

Interrogatives exhibit the following diversity hierarchy: quantity > place > determiner > person > time, thing, manner. Investigation of ontological diversity asymmetries in indefinites and deictics is highly complicated by the cross-cutting categories of indefiniteness and deictic distinctions, and we leave them out of our focus.

The thing and manner interrogatives are the least diverse, showing only phonological developments (cf. *so* 'what' > *ho*; *sar* 'how' > *syr*, *sir*, *sori*, *har*, *hir*). The time interrogative shows phonological developments (e.g. *kana* 'when' > *kan*, *ka*) or borrowing (see Section 20.7), but no indigenous non-

phonological internal diversity. The person interrogative *ko*(*n*) 'who' is supplemented by the former demonstrative *kova* in a few Balkan dialects (Sofia Erli, Yerli, Rumelian Romani), but never completely replaced. The determiner *savo* 'which, what sort of', apart from its phonological developments (e.g. *saw*, *saj*, *so*, *havo*, *haw*), is supplemented by the genitive of the thing interrogative *sosko* 'which' (< 'of what') in some Vlax dialects, and by the place interrogative *kaj* in a few dialects of the Balkans (see also Section 20.4). In Piedmontese Sinti, the expected **(s)avo* is replaced by the form *k-avo*, possibly a reinforcement of **avo* through contamination by *kaj*. Place interrogatives exhibit significant diversity (see Chapter 18 for details).

Quantity is clearly the most internally diverse ontological value with interrogatives. There are four groups of forms. First, forms deriving from Early Romani **keti* (e.g. *keti*, *keci*, *keči*, *kiti*, *kicy*, *kiči*, *kisi*, *kaći*, *gaći*) are retained in the northern and Central dialects, some Balkan dialects (e.g. Prilep Arli, Sepečides, Erli, and Muzikanta), and some Vlax dialects (e.g. Lovari, Xoraxane, Dasikano, and Ukrainian Romani). Second, especially in Balkan and Vlax dialects, there are numerous forms containing the root *-bor* prefixed with various deictic elements (e.g. *a-bor* and *ke-bor* in Rumelian Romani; *ka-bor* in Taikon Kalderaš, Dasikano, and Ajia Varvara; *ko-bor* in Arli of Gilan and Florina, Yerli, Varna Bugurdži, and Rešitare; *ki-bor* in Sofia Erli, Kosovo Bugurdži, Malokonare, Muzikanta, and Gadžikano; *kide-bor* or *kibedor* in Rumungro;[14] and *ta-bbornə* in Abruzzian Romani). Third, some Balkan and Vlax dialects possess forms containing the root **-zom* prefixed with various deictic elements (e.g. *ko-zom* in Arli of Gilan and Florina, Sofia Erli, Yerli, Varna Bugurdži, Zargari, Priština Gurbet, Varna Kalajdži, and Rešitare; *ka-zom* in Kalburdžu and *ka-zum* in Serbian Kalderaš; *ki-zom* in Nange and the Varna dialects; and *ki-rom* in Gadžikano). Fourth, North Vlax dialects possess the form *so-de*, possibly derived from the thing interrogative. Some dialects have up to three quantity interrogatives.

20.7. Borrowing

In this section we discuss borrowing asymmetries in interrogatives and indefinites. Deictics are almost never borrowed (see Chapter 18 for rare exceptions), and so they are left out of the discussion. There is no global ontological hierarchy of borrowing. Instead there are three, partly conflicting, local asymmetries, which are discussed in detail in the text: one concerning the borrowing of interrogatives, one concerning the borrowing of indefinites, and one

concerning the borrowing of indefiniteness markers. Nevertheless, it is still possible to generalise the following partial (and partly overlapping) asymmetries from the three local asymmetries (18)–(19):

(18) Determiner, time > thing, place > person
(19) Determiner, time > place > quantity, cause/goal > manner

Thing and manner **interrogatives** are never borrowed. Person interrogatives in some dialects appear to be borrowed in the nominative: *koj* in Prilep Arli and Sofia Erli from Macedonian and Bulgarian, and *ko* in the Central dialects, Kosovo Bugurdži, Priština Gurbet, and Dasikano from Serbian/Croatian. Serbian Kalderaš and Austrian Lovari show an alternation between the indigenous *kon* and the Serbian/Croatian *ko*. In all instances, the indigenous oblique forms of the person interrogative are retained. The nominative loans show a remarkable phonological similarity to the indigenous form *kon*, and so the process is contact-induced contamination of the indigenous form rather than proper borrowing. Moreover, the form *ko* might, in some instances, be just a result of phonological erosion (**kon* > *ko*).

There are rare instances of borrowing of determiner interrogatives (North Vlax *če* from Rumanian), cause/goal interrogatives (*mīre* in some Lovari varieties from Hungarian), place interrogatives (Abruzzian Romani *kwa* from Italian, South Central *mere* from Hungarian, see Chapter 18), and quantity interrogatives (Kaspičan *kač* from Turkish, Latvian Romani *cik* from Latvian). The quantity interrogative *skaći* in Ukrainian Romani is a contamination of the indigenous *kaći* by East Slavic *s-kol'ko*. Temporal loans are the most frequent: Prilep Arli *ko*(*g*)*a* from Macedonian, Gilan Arli *keda* and Vendic and Slovene Romani *kada* from Serbian/Croatian, Polish, Lithuanian, and Estonian Romani *kedy* (*k'edy*, *kidi*) from Polish, Russian Romani *koli* and Southeast Ukranian *kala* from East Slavic, and Piedmontese Sinti and Abruzzian Romani *kwando* (*kwandə*) from Italian. Unlike the loans of determiner, cause/goal, place, and quantity interrogatives, which supplement indigenous forms, the temporal loans completely replace them. Frequency of borrowing thus renders the following ontological hierarchy: time > place, quantity, cause/goal, determiner > manner, thing, with the position of the person value depending on interpretation. Even though loans of time interrogatives are clearly the most frequent, they are not always implied by borrowing of other interrogatives.

There appear to be few absolute constraints on the borrowing of **indefinite** word-forms. Indefinites of any ontological value may be borrowed, except perhaps for cause/goal indefinites. We have presented the relevant data in

Chapter 19, although from a different perspective. Here we will only provide generalisations concerning ontological asymmetries. Borrowing of manner and quantity indefinites is, irrespective of their indefiniteness value, rare or rarely attested (see Chapter 19). As for the other ontological values (determiner, person, thing, place, and time), the cross-cutting category of indefiniteness plays a significant role. The statistical asymmetries in (20)–(22) are based on a number of dialects that borrow an indefinite of a given ontological category and indefiniteness; the overall statistical asymmetry is given in (23).[15]

(20) Universal: determiner > time > person > thing, place
(21) Negative: time > thing > place > person > determiner
(22) Specific (and other): determiner > time > thing > place > person
(23) Overall: time > determiner > thing > place > person

Some asymmetries hold irrespective of indefiniteness: temporal loans are always more frequent than person, thing, or place loans; and place loans are never more frequent that thing loans. Other asymmetries are contingent on indefiniteness values. First, universal person indefinites are more likely to be borrowed than universal thing and place indefinites, while person indefinites of other indefiniteness values are less likely to be borrowed than the corresponding thing and place indefinites. And second, the determiner is the most contact sensitive value with universal and specific indefinites, but the least contact sensitive value with negative indefinites. Some of the statistical asymmetries appear to be supported by implicational asymmetries (cf. Elšík 2001a).

One ontological asymmetry concerns the borrowing of indefiniteness markers rather than indefinite word-forms. In some dialects, determiners are the only indefinites to borrow a certain indefiniteness marker: e.g. *jek-far* ‘a certain’ in Kosovo Bugurdži (the suffix *-far* is only found in the determiner), *hidž-ekh* in Sepečides (the other negative indefinites are based on the determiner), *ni-jek* ‘no, none’ in Prilep Arli (the other negative forms are word-form loans). Further, dialects that borrow an indefiniteness marker in some but not all ontological categories will usually have it in the determiner: e.g. Kumanovo Arli *ne-kori* ‘somewhere’ as well as *ne-savo* ‘some’, or Serbian Kalderaš *ni-sar* ‘in no manner’ as well as *ni-sosko* ‘no, none’. Thus, determiners appear to be more prone to borrowing of indefiniteness markers than other ontological values.

Chapter 21
Lexicality

We define *Lexicality* as the transparency of lexical meaning and conceptual symbolism. An item that is high on the lexicality scale represents a more stable and independent and thus more transparent concept or object of reality, while items that are low on the lexicality scale show stronger context-dependency of their meaning. We distinguish two (sub)categories. The category of *auxiliarity* is applicable to verbal and substantival word classes, differentiating, respectively, the copula/existential verb and pronouns (the auxiliaries) from lexical verbs and lexical nouns (the non-auxiliaries). The category of *nominal lexicality* differentiates substantivals from adjectivals and, within the latter, lexical adjectivals (descriptive adjectives) from operators (e.g. demonstratives, articles, and pronominal possessives). Substantivals are more lexical than adjectivals in that they encode more stable concepts than adjectivals do.

Auxiliaries are more differentiated, while non-auxiliaries are more complex, likely to extend, and more likely to be borrowed. Items of greater nominal lexicality are more complex, more differentiated, more likely to extend, and more likely to be borrowed than items of lesser nominal lexicality.

21.1. Complexity

One of the participial markers which Early Romani inherits from the MIA inventory is *-(*i*)*na* > *-in*. It serves in some dialects as a marker, or an extended marker, of the past tense of some verbs. Overall, a closed set of lexical verbs is more likely to show *-in-* than the copula, and where *-in-* appears in the copula, it will also appear at least in a selection of lexical verbs (lexical verbs > copula).

The widest distribution is with a small group of psych-verbs ending in *-a*, as in *asa-* 'to laugh',[1] as well as with the two monoconsonantal verb roots *d-* 'to give' and *l-* 'to take'. With psych verbs, *-in-* may be the principal perfective marker (Šóka Rumungro *asa-ň-om* 'I laughed'), though more commonly it acts as an intrusive marker and is followed either by the respective perfective marker for stems in *-n-* (Lithuanian Romani *asa-n-dj-om* 'I laughed'), or by the middle/intransitive/unaccusative perfective marker *-il-* (Crimean Romani

asa-n'-il'-om 'I laughed'), or by a combination of both these markers (Florina Arli *asa-n-d-il-om* 'I laughed'). With *d-* 'to give' and *l-* 'to take', *-in-* may similarly serve as the sole perfective marker (Yerli *d-in-om* 'I gave'), or it may be followed by the perfective marker for the *n*-class (Slovene Romani *d-in-dž-om* 'I gave'). The alternative to the extension in *-in-* is either a zero-marker in the perfective (Lovari *d-em* 'I gave'), or a palatal extension (Kalderaš *d-ij-em* 'I gave'), with the monoconsonantal stems, or the unaccusative extension *-il-* with psych verbs (Lovari *asa-jl-em* 'I laughed'). We may therefore view the extension in *-in-* as an extension of the perfective stem, and one that contributes to it complexity.

The instrusion in *-in-* may also appear in the copula, itself a monconsonantal stem in *s-* or *h-*. Its appearance in the copula is constrained by person (third person vs. other persons) and tense (present vs. past), whereas with lexical verbs the extension may be constrained by the type of perfective marker: personal (jotated), or adjectival (non-jotated). Table 21.1 summarises the types of distribution of the perfective stem extension *-in-*.

Type 1 dialects are common in the south-central Balkans (Arli of Prilep and Kumanovo, Sofia Erli, Yerli, Varna Bugurdži, Sepečides), but include Slovene Romani as well. Here, the intrusion is found with monoconsonantal verbs and psych verbs in *-a* (Kumanovo Arli *d-in-g-jum* 'I gave', *dara-n-d-il-jum* 'I feared'), with the third-person copula in the past (Kumanovo Arli *ov ine* 'he was', cf. < **h-in-e*; cf. present *ov i* 'he is'), and with the first- and second-person copula (Kumanovo Arli *in-jum* 'I am'). Type 2 dialects are similar, but lack *-in-* in the first- and second-person copula. This pattern is common in

Table 21.1. Intrusion -in- as stem extension

	d- / *l-* , *asa-* (all)	*d-* / *l-* , *asa-* (non-jot.)	Other verbs (non-jot.)	COP.3 PAST	COP.3 PRES	COP.1–2
Type 1	+		–	+	–	+
Type 2	+		–	+	–	–
Type 3	+		–	–	+	–
Type 4	+		–	–	–	–
Type 5	–	+	+	–	–	–
Type 6	–	–	–	–	–	–

the southwestern Balkans (Arli of Skopje, Florina and Karditsa, Epiros), but is also found in Sípos and Nógrád Rumungro as well as in Nange (cf. Karditsa *asa-n-d-il-om* 'I laughed', *d-in-om* 'I gave', *is-in-es* 's/he was', but *som* 'I am'). Type 3 comprises East Slovak and Finnish Romani. Here, *-in-* appears in the copula only in the third-person present (*hin* 's/he is'). In Type 4 (Crimean, Bohemian, Rumelian, Ukrainian, and Klenovec Rumungro), *-in-* does not appear in the copula at all. The Northeastern dialects constitute Type 5, where only the adjectival formants of the perfective conjugation show extension in *-in-* (thus Polish Romani *d-yj-om* 'I gave' but *d-yn-e* 'they gave'), but this extension is diffused to other verbs as well (Polish Romani *xa-n-e* 'they ate', *g-yn-e* 'they went', *mukh-n-e* 'they left'). Some Vlax dialects partly agree (Lovari *d-em* 'I gave' but *d-in-e* 'they gave'). Individual spreads of the extension to other verbs are also found in various other dialects. Frequently affected are verbs whose stems end in *-d-*, by analogy to *d-* 'to give' (Lovari, Varna Bugurdži *trad-in-e* 'they drove', *bold-in-e* 'they turned over'), or unaccusative verbs of motion and state (Yerli *ušt-in-e* 'they stood up', Gadžikano *ač-in-e* 'they stayed'). Absence of *-in-* in Type 6 is a feature of the Sinti group, most Vlax dialects, and some of the northern Bulgarian dialects.

Relative segmental complexity is a feature of the inflection of attributive demonstratives, compared to attributive adjectives (demonstratives > adjectives). While oikoclitic adjective inflection for gender, number and case is carried by vowels only (M.SG.NOM *-o*, M.SG.OBL *-e*, F.SG.NOM *-i*, F.SG.OBL *-a*/*-e*, etc.), in demonstratives, the inflectional suffix is usually composed of a consonant and a vowel (M.SG.NOM *-va*, M.SG.OBL *-le*, F.SG.NOM *-ja*, F.SG.OBL *-la*/*-le*, etc.). Exceptions are found in individual dialects, such as Lovari (M.SG *kad-o*, F.SG *kad-i*, but PL *kad-ala*), where singular demonstratives adopt adjectival inflections.

21.2. Differentiation

Auxiliaries (the copula and pronouns) tend to be more differentiated than lexical verbs and nouns. The copula shows a greater differentiation in terms of TAM distinctions, inflectional irregularity, and through its greater propensity to co-occur with subject clitics. Pronouns are more differentiated than nouns in that they are more likely to retain synthetic case inflection; in that they exhibit a greater stem differentiation; and in that they are more likely to have irregular and differentiated genitive (possessive) forms. On the other hand, pronouns are less likely to retain gender distinctions than nouns. In the cat-

egory of nominal lexicality, substantivals are more differentiated than adjectivals; and descriptive adjectives and cardinal numerals tend to be more differentiated than less lexical modifiers. We now discuss the relevant phenomena and developments in more detail.

The **copula** is more differentiated than lexical verbs in terms of TAM distinctions (see Chapter 13). The pluperfect and the imperative values are distinctly encoded both in the copula and in lexical verbs. Also both in the copula and in verbs, there are two sets of forms that encode different internal distinctions in the domain of indicative past and real/potential conditional (see Table 13.2 and the accompanying discussion in Chapter 13). The greater differentiation of the copula stems from the fact that its subjunctive and also future forms are based on distinct non-indicative roots (see below). Consider forms of lexical verbs, exemplified by the verb *ker-* 'do, make', and of the copula in the subjunctive-present-future domain (Table 21.2; all forms are first-person singular).[2]

In dialects of Type A (e.g. those of Central-East Europe and South Vlax), there is a single form for the (present) subjunctive and the (indicative) present in lexical verbs. However, the two functions are distinctly encoded in the copula: the present is based on the indicative root *s-*, while the subjunctive is based on a non-indicative root (*ov-* or *av-*, see below). In dialects of Type B (e.g. Finnish Romani, Piedmontese Sinti, and Taikon Kalderaš), which appear to preserve the Early Romani state, there is a single form for the present and the future. Again, the two functions are distinctly encoded in the copula. Only in dialects of Type C (e.g. numerous Balkan dialects) is there an identical number of TAM distinctions in lexical verbs and in the copula in the subjunctive-present-future domain. Moreover, most dialects possess a distinct past subjunctive form, which is different from the indicative past (e.g. *uliljom* '[that] I were' vs. *somas* 'I was'). There is no distinction between the subjunctive and the indicative in the past forms of lexical verbs.

Table 21.2. Patterns of TAM differentiation in lexical verbs and in the copula

	SUBJ	PRES	FUT	SUBJ	PRES	FUT
Type A1	*kerav*		*kerav-a*	*ovav*	*som*	*ovav-a*
Type A2	*kerav*		*ka kerav*	*ovav*	*som*	*ka ovav*
Type B	*kerav*	*kerav-a*		*ovav*	*som*	*ovav-a*
Type C	*kerav*	*kerav-a*	*ka kerav*	*ovav*	*som*	*ka ovav*

The copula is also the most differentiated verb in terms of number of its roots and the irregularity of their formal relations. The Early Romani copula possessed four suppletive roots: the indicative *s-* and *h-*, the non-perfective non-indicative *ov-*, and the perfective indicative *u-*. In addition, a suppletive third-person negative form of the copula is found in all dialects; the synchronic suppletion probably developed already by the Early Romani period. Individual dialects either redistribute the indicative roots according to the categories of person and tense (see Chapter 7 for further discussion), or they generalise one of them, thus reducing the degree of differentiation in the copula. In some dialects, some or all non-indicative forms of the copula have been renewed through grammaticalisation of the verb *av-* 'come' and integration of some of its forms into the copula paradigm. The renewal may affect only the non-perfective non-indicative root, so that the non-indicative roots of the copula are *av-* and *u-* (e.g. in some varieties of Slovak Romani), or it may affect both non-indicative roots. If the latter, then there is a reduction in the number of copula roots, since the perfective stem of the verb 'come' is derived from the non-perfective stem by suffixation of a perfective marker (e.g. *av-* > *av-il-*). The number of suppletive copula roots in different dialects thus varies from five (as in Early Romani) to three (e.g. indicative *is-*, non-indicative *jav-*, and negative *-ne* in Russian Romani). The lexical verbs, on the other hand, possess only two stems, the non-perfective and the perfective. They are almost never suppletive and, with the exception of a few verbs, the perfective stem derives from the non-perfective one in a regular way.

One more phenomenon attests to the greater differentiation of copula with regard to lexical verbs: the greater likelihood of the co-occurrence of clitic subject pronouns with the copula. Early Romani appears to have possessed two sets of third-person nominative pronouns, one in *l-* (cf. *l-o* 'he', *l-i* or *l-a* 'she', *l-e* 'they'), which is cognate with the oblique set, and one formed from the demonstrative set in *o-* (cf. *ov* 'he' < **o-va*, *oj* 'she' < **o-ja*, *ol*/*on* 'they' < **ola*/*ona*). The latter are likely to have served as emphatic pronouns initially, but have gradually taken over the role of default anaphora in almost all dialects. Consequently, the set in *l-* has retreated. It appears only in enclitic position (except in Roman, where it can also stand pre-verbally), and in most dialects it is confined to copulas or even to non-verbal predications: presentatives (e.g. Dasikano *eta-lo* or Prilep Arli *ake ta-lo* 'there he is') and place deictics (e.g. Austrian Lovari *kate-lo* 'here he is'), and/or interrogatives (e.g. Lovari *kaj-lo?* 'where is he?'). In some dialects, the clitics have reduced to vocalic suffixes (e.g. *lo* > *-o*). Table 21.3 shows the distribution patterns of the subject

clitics or suffixes; we have disregarded tense and transitivity distinctions here (see Chapters 13 and 15).

The figure shows that subject clitics may occur independently with verbal predications, and with non-verbal predications. In Type A (which is typical of the Balkans, e.g. Florina Arli, Yerli, Varna Bugurdži, Dasikano, Ajia Varvara, and Rešitare) and Type B (attested in Polish Lovari), they are present with non-verbal but not with verbal predications; and in Types F (e.g. Sinti and some South Central dialects, viz. Roman and Klenovec Rumungro) and G (e.g. Finnish and Lithuanian Romani, most Central dialects, Prizren Arli, Kalderaš, and Rakarengo), they are present with verbal but not with non-verbal predications. Moreover, there is no apparent link between the individual types of non-verbal predication, and clitics may follow independently either presentative or local deictic constructions (Type A) or interrogatives (Types B and C; the latter attested in Slovene Romani). Nevertheless, within the verbal predications, the presence of clitics with lexical verbs implies their presence with the copula. In Types C, D (attested in Prilep Arli), and G only the copula may co-occur with the subject clitics, while in Types E (attested in Austrian Lovari) and F the clitics co-occur with at least some lexical verbs, too.

A cross-dialectally marginal, but outstanding, instance of a greater differentiation of **pronouns** is found in Italian Sinti and the Apennine dialects. Due to prolonged contact with Italian there has been a profound reduction of the category of inflectional case in these dialects, which is more likely to affect nouns than pronouns (cf. Elšík 2000a: 3).[3] In Piedmontese and Lombardian Sinti, nouns have lost case inflection, while pronouns retain the whole

Table 21.3. Distribution of subject clitics

	Presentative	Interrogative	Copula	Lexical verb
Type A	+	–	–	–
Type B	–	+	–	–
Type C	–	+	+	–
Type D	+	+	+	–
Type E	+	+	+	+
Type F	–	–	+	+
Type G	–	–	+	–

range of Early Romani cases. In Abruzzian Romani, nouns have likewise lost case inflection, while pronouns retain all cases but the ablative. Calabrian Romani has now lost case even in personal pronouns: the original nominative has been generalised in the singular pronouns, and the original locative in the plural pronouns (e.g. *lamen-də* 'we').[4] In all of these dialects, there are only lexically restricted remnants of the original non-nominative cases with nouns. For example, the locative is found with names of localities in Piedmontese Sinti (e.g. *Milanate* 'in Milan') and may be synchronically considered to be a de-nominal adverb; the sociative is a means to derive de-nominal adjectives in Abruzzian Romani (e.g. *xoljinas* 'angry' < *xoli* 'anger'); and the genitive is only retained in agentive derivations (e.g. *maseskero* 'butcher' < *mas* 'meat').

Personal pronouns of the first and the second persons (and, in some dialects, also reflexive pronouns) show greater differentiation of stems than other substantivals. The latter construct all of their forms on the basis of two stems, the base (nominative) stem and the oblique stem; they construct the genitive by means of suffixing a genitive marker to the oblique stem (e.g. oblique *rom-es-* 'husband' > genitive *rom-es-ker-*). The first- and second-person pronouns, in addition, possess a specific genitive stem which is not based on the oblique stem (e.g. oblique *am-en-* 'we' vs. genitive *am-ar-* 'our'). Moreover, the genitive marker in these pronouns is mostly irregular (see also Section 21.3). Also, personal pronouns possess two or more genitive variants in numerous dialects (e.g. Gurvari *mūro* vs. *muro* 'my' or Gilan Arli *kl-* and *t-* 'your.SG'; cf. Elšík 2000a: 16–18), whose distribution is determined by various syntactic factors; such variation is much rarer with nouns (cf. the genitive variants *-k-*/*-g-* vs. *-kīr-*/*-gīr-* in Latvian Romani).

Nouns, on the other hand, are more likely to retain gender distinctions than pronouns. Most dialects that have recently lost the gender distinction in the nominative of the third-person pronouns due to convergence with genderless contact languages (see Section 8.3 for details) still retain gender fully intact in nouns, as evidenced by their own inflection (as well as by the inflection of their modifiers). Only some modern varieties of the non-native Finnish Romani show gender dissolution also in noun inflection: original masculines occasionally take feminine oblique markers and vice versa (e.g. *čaj-es-* 'Gypsy girl' with the masculine suffix *-es-*).

As for **nominal lexicality**, substantivals are more differentiated than adjectivals with respect to the category of case (see Chapter 16): the former inflect for the eight-value substantival case, while the latter, when they are employed in their primary function as (preposed) modifiers in noun phrases, inflect for

the two-value adjectival case in most dialects (e.g. *bar-e džukl-eske* 'to a big dog'). In a few dialects, especially in some modern varieties of Sinti, adjectival modifiers have lost case inflections altogether, even if case is retained with substantivals (e.g. *baro džukl-eske*). In Slovene Romani and in some varieties of Russian and Slovak Romani, preposed adjectival modifiers inflect for the substantival case due to convergence with Slavic languages (e.g. *bar-eske džukl-eske*). The full case agreement neither conforms to, nor violates, the above asymmetry.

Adjectivals are substantivised and inflected for the substantival case, when they are postposed (e.g. *džukl-eske bar-eske*) or when they are employed as heads of noun phrases (e.g. *bar-eske* 'to a big one'). The definite article and, where they exist, also non-emphatic variants of demonstratives and pronominal genitives (possessives) cannot be substantivised, and so their number of forms and their degree of differentiation is greatly reduced. They can also be indeclinable: the article does not inflect at all, for example, in Yerli; indeclinable demonstrative variants are found in many dialects, including Lithuanian Romani (e.g. *da* 'this'); and pronominal genitives are indeclinable in Iranian Romani (e.g. *mi* 'my') and some Sinti varieties (e.g. *mur* 'my').[5] On the other hand, the major class of lexical adjectives (viz. the vocalic class) is always inflected in modifier position, at least for number and gender. In some dialects (e.g. Taikon Kalderaš and Razgrad Drindari), most adjectivals encode gender in the oblique, while there is gender neutralisation in pronominal genitives (e.g. masculine *koř-e* vs. feminine *koř-a* 'blind', but gender-indifferent *myř-e* 'my' in Taikon). Thus, the more lexical descriptive adjectives tend to be more differentiated than less lexical adjectivals.

Another instance of a lesser differentiation of less lexical modifiers is found in expressions of cardinal quantity. While cardinal numerals, at least the lower ones (see Section 11.2 for details), show case agreement, interrogative, indefinite and deictic quantity pro-words are always uninflected (e.g. *trin džene* 'three people' and *trin-e dženenca* 'with three people' vs. *keti džene* 'how many people' and *keti dženenca* 'with how many people').

21.3. Extension

More lexical forms, especially inflections, extend to less lexical words. This direction of extension represents regularisation of the less numerous auxiliaries and operators, their formal assimilation to the more numerous lexical words. We only discuss one salient instance in this section.

In a few dialects of different dialect groups, forms parallel to regular nominal genitives have been created in the first- and second-person pronouns. Like other regular genitives, they are derived from the oblique stem by a regular genitive marker. In some dialects, these innovative forms are limited to a certain number: to the singular in Serbian Kalderaš (1SG *man-g-*, 2SG *tu-k-*), and to the plural in Lombardian Sinti, Abruzzian Romani, and the Istrian variety of Slovene Romani (e.g. 1PL *men-gr-*, 2PL *tumen-gr-*). The regularised genitives seem to have replaced the original genitive forms in all of their functions in Abruzzian Romani and some modern varieties of Finnish Romani. In Serbian Kalderaš and Rumelian Romani, however, they only exist in a construction with the preposition *bi* 'without', which is the only preposition that may govern the genitive with nouns (e.g. Serbian Kalderaš *bi mango* 'without me').[6]

21.4. Borrowing

Lexicality favours borrowing in two major word classes, verbs and nouns. Romani dialects borrow lexical verbs, but seldom either copula or other transition or existential verbs (such as 'to become', or 'to have'). On the other hand, modal verbs are often borrowed, though we have no attestation of a dialect that borrows modals but not lexical content verbs. We might therefore postulate the hierarchy for borrowing: content verbs > modal verbs > existential verbs. In nominals, Romani borrows numerous nouns, but there is no attestation of the borrowing of either demonstratives or of complete forms of pronouns (exceptions are plurality markers on third-person plural pronouns in some dialects, and the adjusted form *min* 'I' in Zargari; cf. Elšík 2000a). For nominals we may thus postulate: content noun > pronoun.

The reverse relation holds for the distribution of borrowed person concord on verbs. On the whole, lexical content verbs tend not to show any borrowing of person concord markers. There are several exceptions to this. Slovene Romani borrows concord markers from Croatian, which are used throughout the verb system, i.e. also with inherited (pre-European) roots. Several dialects use Greek-derived *-i* in the third-person singular of borrowed verbs, or sometimes also with selected pre-European verb roots (in Slovene Romani, it is used with all verbs). Turkish verb inflection is used with either all, or with a sub-set of Turkish-derived verbs in many of the dialects of the Balkans, as well as in Zargari (Azeri inflection with Azeri verbs). Russian Romani tends to show Russian inflection with many Russian verbs, and Epiros and Dendropotamos Romani have Greek concord markers with a selection of borrowed verbs (from

Greek, and in Dendropotamos also from Turkish). The use of source-language inflection is even more widespread in modal verbs, however. First, some dialects that do not show borrowed inflection with borrowed lexical verbs may show it with modal verbs. Thus Kosovo Bugurdži *moram, moraš, mora* 'I, you, s/he must' (Serbian), Rešitare *trjabva*, past *trjabvaše* 'must' (Bulgarian). Second, dialects that show borrowed inflection even with just some lexical verbs are likely to show it also with modal verbs: Epiros *oleski phen irizi* 'his sister is returning' (Greek 3SG *-i*), but also *borume te džas* 'we can go' (Greek 1PL *-ume*); also Kaspičan (1), where the verb 'began' is marked for the Turkish 3SG.PAST and 'cry' for the Turkish 3SG.SUBJ:

(1) *Mi phen bašladə te baarsən kana tharde amare kera.*
my sister began to cry when they-burned our houses
'My sister started to cry when they burned down our houses.'

The conclusion appears to be that the adoption of full verb inflection is licensed by the adoption of verb inflection with borrowed modals (Mod-INFL > Lex-INFL). This is not, strictly speaking, a reversal of the lexicality hierarchy depicted above with respect to the borrowing of the verbal root itself, but a lexicality-related condition on the borrowability of inflectional material.

Chapter 22
Associativity

The category of associativity is marginal in its cross-dialectal distribution, being attested only in some Rumungro varieties, where it is calqued on Hungarian. Following Corbett (2000: 101–111), we consider associativity to be a category distinct from number. In Rumungro, the category of associativity combines with the plural: there is a regular (non-associative) plural and an associative plural. Encoding of associativity is restricted to names of persons and nouns denoting professions. The associative plural of a noun denotes the referent of the corresponding singular together with a group of associated persons (e.g. *Alenangere* 'Alena and her family', *lakatošingere* 'the locksmith and his work group'), rather than plurality of referents denoted by the singular. The latter is expressed by a regular plural (e.g. *Aleni* 'Alena's', *lakatošša* 'locksmiths').

Associative forms in Rumungro are more complex than the corresponding non-associative forms. Consider the nominative, oblique, and genitive forms of the noun *lakatoši* 'locksmith' in Šóka Rumungro (Table 22.1).

The singular oblique marker *-is-* of this noun may be segmented into a classification suffix *-i-* (which is shared by nouns of a single inflectional class) and the masculine singular oblique suffix *-s-* (which is shared by all masculines). Similarly, the regular plural oblique marker *-en-* consists of the classification suffix *-e-* and the plural oblique suffix *-n-*. The associative forms inflect almost like substantivised genitive forms of the regular plural. The significant difference is in the quality of the classification suffix. The classification suffix of the associative plural is that of the singular (*-i-*), not of the regular plural (*-e-*). On the other hand, the associative is clearly marked as non-singular by the plural oblique suffix *-n-*. The functions of markers involved in the formation of the associative forms may be summarised as follows: The oblique marker *-n-* indi-

Table 22.1. Associative forms in Rumungro ('locksmith')

	SG	PL	PL.ASS
NOM	*lakatoš-i*	*lakatošš-a*	*lakatoš-i-n-ger-e*
OBL	*lakatoš-i-s-*	*lakatoš-e-n-*	*lakatoš-i-n-ger-e-n-*

cates non-singular reference: accordingly, it occurs in the regular plural and in the associative plural, but not in the singular. The classification marker *-i-* indicates that a single referent is in focus: accordingly, it occurs in the singular and in the associative plural, but not in the regular plural. The associative forms are clearly more complex than the corresponding non-associative forms, as they are in fact substantivisations of genitive forms in *-C-n-ger-* (where *-C-* represents a classification suffix).[1]The associative forms are also more differentiated as to inflectional class, since the classification suffixes in the associative plural (cf. *-o-*, *-i-*, *-u-*, *-a-*) are more diverse than those in the regular plural (cf. only *-e-* or variable *-e-* ~ *-o-*). There is no differentiation asymmetry in the cross-cutting category of case. The other criteria are either irrelevant (e.g. diversity, since our sample contains a single independent instance of encoding of the category), or they do not render any obvious asymmetries.

Chapter 23
Chronological compartmentalisation

This last chapter of Part II (Data Evaluation) deals with a distinction that can be subsumed under the label of a grammatical category only with difficulties. Nevertheless, the distinction lends itself to the same set of criteria that have proved to be relevant for recognising asymmetries in the grammatical categories we have examined.

Chronological compartmentalisation is our term for the morphological encoding of the origin of lexemes. Since the Early Romani period, major parts of the Romani lexicon have been divided into two compartments: the "indigenous" compartment, which consists of lexemes inherited from Indo-Aryan plus Asian (West Iranian, Armenian, Kurdish) and early Greek borrowings, and the "borrowed" compartment, which consists of later Greek and post-Greek borrowings. Importantly, this compartmentalisation is not based on specialists' knowledge about the origin of the lexemes. Instead, it is directly encoded in the structure of the language and, at least in principle, accessible to its speakers.

The terms "indigenous" and "borrowed" compartment are inaccurate and potentially misleading in three respects. First, the "indigenous" compartment also contains early borrowings into Romani. Second, individual indigenous lexemes may behave like those of the "borrowed" compartment and, vice versa, individual borrowed lexemes may behave like those of the "indigenous" compartment. And, third, some classes of (indigenous or borrowed) lexemes are not subject to chronological compartmentalisation in the above sense, i.e. they are located outside of either compartment, and so irrelevant for the distinction discussed here. For this reason, it is advantageous to employ a different terminology. In order to avoid the ambiguity of the terms *thematic* and *athematic*, which have been widely used in Romani linguistics (cf. especially Bakker 1997), we introduce the terms *oikoclitic* and *xenoclitic* for the "indigenous" and "borrowed" compartments, respectively. The terms are used to designate the compartments but also, metonymically, word classes that fall within these compartments and markers that are associated with these word classes. Oikoclitic and xenoclitic are the values of the distinction of chronological compartmentalisation.

Seeds of chronological compartmentalisation were introduced in the Early Romani period during intensive contact with Greek, when a great number

of Greek grammatical suffixes were borrowed together with Greek lexicon. However, it was the contact with post-Greek contact languages of Romani that brought chronological compartmentalisation to life. By then, still in the Early Romani period, numerous Greek grammatical markers had been abstracted from their source lexemes and applied to the new loanwords as well, thus signalling their foreign origin in Romani and constituting the xenoclitic compartment. Pre-Greek lexemes and a few early Greek loans, on the other hand, largely retained their indigenous morphology, constituting the oikoclitic compartment.

Although the Early Romani compartmentalisation has been, on the whole, preserved in all dialects, it has been open to further developments. First, there has been structural interaction between the oikoclitic and the xenoclitic compartments in various structural domains (see Section 23.3 on Extensions for details). Second, some of the Greek-derived markers were later, in dialect-specific developments, replaced or supplemented by markers borrowed from post-Greek contact languages,[1] some of which have diffused backwards in time within the xenoclitic compartment: for example, the Rumanian-derived plural suffix *-uri* in Vlax applies not only to Rumanian and post-Rumanian loans but also to pre-Rumanian (including Greek) loans. Third, new structural domains have been drawn into the distinction of chronological compartmentalisation through re-iteration of the contact process described above for Greek: for example, it appears that xenoclitic derivation of female nouns came about, in numerous dialects, due to contact with South Slavic. To conclude: although the substance (the markers) and structural details (the domains) of chronological compartmentalisation have been changing, the general pattern has remained intact since Early Romani.

Table 23.1 presents the structural domains of compartmentalisation found in current Romani dialects, and gives examples of oikoclitic and xenoclitic markers in each domain. Greek-derived xenoclitic markers are preceded by "G"; later xenoclitic markers are separated by a semicolon.

As is evident from Table 23.1, the structural domains of chronological compartmentalisation are both inflectional and derivational. As for inflection, the compartmentalisation is best developed in nouns, where there are a number of xenoclitic inflectional classes in any dialect, and in adjectives. In verbs, the compartmentalisation is restricted to a single person–number inflection and to participle suffixes. Derivational compartmentalisation is well developed in numerous nominal domains: diminutives, female nouns (i.e. derivations of female counterparts from nouns denoting persons), agentives (both de-nominal and de-verbal), nominalisations (de-adjectival and de-verbal abstract

Table 23.1. Domains and selected markers of chronological compartmentalisation

Domain		Oikoclitic	Xenoclitic
Inflection	Nouns	e.g. *-ó, -í*	e.g. G *-os, -is* ~ *-i, -a*
	Adjectives	e.g. *-ó*, PL *-é*	e.g. G *-o*, PL *-a*
	Verbs: 3SG	*-(e)l*	G *-i*
	Verbs: participle	*-d-, -l-* etc.	G *-men*
Derivation	Diminutives	*-oř-*	G *-ak-*; *-ičk-* etc.
	Female nouns	class, *-n-*	*-kinj-, -ink-* etc.
	Agentives	(genitives)	*-ar-, -tor-, -oš-* etc.
	Nominalisations	*-ben* ~ *-pen*	G *-mos; -šag-* etc.
	Adjectives (< nouns)	*-al-, -an-* etc.	G *-ik-, -itik-*
	Adverbs (< adjectives)	*-es*	G *-a, -on-*

nouns), de-substantival adjectives, and de-adjectival adverbs. Although chronological compartmentalisation is, at least in some dialects, also relevant for verb derivation (de-verbal causatives, de-substantival verbs, and factitives and inchoatives, i.e. de-adjectival causatives and middles), we do not have enough comparative data here, and do not include this domain in our discussion. Finally, there is also a class of xenoclitic elements used as non-inflectional adaptation markers (these will be discussed in more detail in Section 23.1).

The oikoclitic value tends to be more differentiated and is more likely to extend. The xenoclitic value tends to be more complex, is more diverse, and it is the one that is borrowed; it also appears to be more likely to be exposed.

23.1. Complexity

Xenoclitic elements tend to be more complex than oikoclitic elements. Inflectional and derivational xenoclitic markers mostly show equal morphological complexity as their oikoclitic counterparts, although they tend to be longer (e.g. in participles or female nouns). It is the fact that borrowed lexemes may be adapted through overt *adaptation markers* that clearly makes the xenoclitic value more complex.

Two strategies are available for morphological adaptation of borrowed lexemes of the major word classes (verbs, nouns, adjectives): adaptation through inflectional integration alone, and adaptation through overt adaptation markers as well as inflectional integration. Overtly adapted borrowed lexemes are

more complex than indigenous lexemes of the same word class, while lexemes adapted by inflections alone show equal complexity as indigenous lexemes. Overt adaptation (e.g. by active *-Vn-, -Vz-, -Vsar-*, depending on dialect) generally occurs with borrowed verbs (e.g. *piš-in-, piš-iz-, piš-isar-* 'write' < Slavic *piš-*). Borrowed nouns, on the other hand, are generally adapted by inflections alone. Borrowed adjectives may be overtly adapted in some dialects, while in other dialects they are only adapted by inflectional integration. The overt adaptation of borrowed adjectives is usually restricted to a certain layer of loans: to Hungarian loans in the Central dialects, which are adapted by *-n-* or *-av-* (e.g. Rumungro *kīk-n-o* 'blue' < *kék*, *sirk-av-o* 'grey' < *szürke*); and to Turkish loans in some dialects of the Balkans, which are adapted by *-i* or *-s* (e.g. Sepečides *temiz-i* 'clean' < *temiz* and *dovru-s* 'true, right' < *doğru*).[2]

23.2. Differentiation

The inflection of oikoclitic nouns and adjectives tends to be more differentiated than inflection of xenoclitic nouns and adjectives. We have discussed the nominative–accusative homonymy in xenoclitic nouns in Section 6.2 (see Tables 6.11–12 and the accompanying discussion). An outstanding feature of the xenoclitic inflection of adjectives is gender neutralisation in the nominative, which does not occur in the major class of oikoclitic adjectives (e.g. the xenoclitic *šarg-o* 'yellow.M/F' vs. oikoclitic *kal-o* 'black.M' and *kaľ-i* 'black.F' in East Slovak Romani).

Oikoclitic participles inflect for adjectival categories (number, gender, case), and hence they are more differentiated than the xenoclitic participles in *-men*, which are indeclinable in most dialects. The xenoclitic participles inflect only in a few dialects, viz. Arli and Vendic, where they show the same amount of differentiation as the oikoclitic participles. After undergoing an erosion of the xenoclitic participle suffix **-men* > *-me*, these dialects have re-analysed the suffix as containing the adjectival plural inflection *-e* and, by analogy, created its other inflectional forms (e.g. nominative masculine singular *-m-o*).

23.3. Extension

Extensions occur in both directions: extensions of oikoclitic markers to indigenous lexicon (oikoclitic extensions) and extensions of xenoclitic markers to borrowed lexicon (xenoclitic extensions). However, it appears that oikoclitic

extensions are more prominent than xenoclitic extensions for reasons to be explained below. Table 23.2 summarises the various extensions according to structural domain.

In Table 23.2 and in the following discussion, we distinguish between two types of extension: partial extensions affect only part of the opposite chronological compartment (i.e. a few individual lexemes or a well-defined class of lexemes), while complete extensions affect the whole opposite compartment. Thus, complete extensions result in a loss of chronological compartmentalisation in the relevant structural domain. For example, an extension of the indigenous diminutive marker *-oř-* to all bases, including all borrowed bases, is an instance of a complete oikoclitic extension and results in a loss of chronological compartmentalisation in the domain of diminutives. The extension is termed oikoclitic, since the marker *-oř-* used to be oikoclitic before it underwent the complete extension; synchronically, the marker does not participate in chronological compartmentalisation any more (see below for the concrete data).

Table 23.2 shows that partial extensions frequently take both directions. If only one direction occurs, then it is mostly the xenoclitic extension (in female nouns and agentives). This may be interpreted as a result of a greater productivity of the newer xenoclitic classes in derivation. Complete extensions, on the other hand, either take both directions (in participles and abstract nominalisations), or they are oikoclitic (in adjective and finite verb inflection[3] and in diminutives). The complete oikoclitic extensions in inflection (and also the partial oikoclitic extension in noun inflection) defy an explanation in terms of

Table 23.2. Summary of extensions in chronological compartmentalisation

Domain		Partial	Complete
Inflection	Nouns	both (oikoclitic)	–
	Adjectives	(xenoclitic)	oikoclitic
	Verbs: 3SG	both	oikoclitic (xenoclitic)
	Verbs: participle	both	both
Derivation	Diminutives	both	oikoclitic
	Female nouns	xenoclitic	–
	Agentives	xenoclitic	–
	Nominalisations	both	both
	Adjectives (< nouns)	both	–
	Adverbs (< adjectives)	both	–

productivity, and they will be considered to represent the prominent direction of extension. We now proceed to a more detailed discussion of extensions in individual structural domains.

All dialects retain a general distinction between oikoclitic and xenoclitic inflectional classes in nouns, i.e. there is no complete extension in either direction. Extensions are partial, affecting only selected inflectional classes; they are mostly oikoclitic. In Core Sinti, borrowed masculines in *-o* inflect as indigenous masculines in *-o*, having completely assimilated to their oikoclitic inflection.[4] In Welsh Romani and the South Central dialects, a distinction between oikoclitic and xenoclitic masculines in *-o* is retained, although there is extension of individual inflections. Welsh Romani extends some oikoclitic inflections to the xenoclitic class (e.g. the oblique singular *-es-* rather than **-os-*), while there are extensions in both directions in Roman and western Rumungro. These bidirectional extensions tend toward re-structuring the original oikoclitic vs. xenoclitic distinction into a distinction based on animacy, with animates showing more oikoclitic inflections and inanimates showing more xenoclitic inflections (for a fuller discussion see Elšík 2001b). Unidirectional xenoclitic extensions are rare. Apart from class transfers of individual indigenous lexemes (e.g. the Indic-derived *lindr-i* 'sleep' to the xenoclitic *lindr-a* in numerous dialects), only weak oikoclitic classes are attested as assuming xenoclitic inflection. This is the case, for example, with the indigenous masculines in *-i* in Rumungro (a class consisting of one to two lexemes, depending on variety), which assimilate to the inflection of xenoclitic masculines in *-i*.

Complete oikoclitic extension in adjective inflection is typical of the South Central dialects. In Rumungro, all borrowed adjectives inflect like indigenous adjectives of the vocalic class. In Roman, this holds for Croatian and Hungarian loans, while German adjectives are uniflected as modifiers. The loss of the Greek-derived xenoclitic inflection of borrowed adjectives through complete oikoclitic extension may be reconstructed to have taken place also in Welsh Romani and Sinti. In some of these dialects, however, new xenoclitic classes of adjectives have been created. Thus in Welsh Romani, English and Welsh loans as well as adjectives derived by xenoclitic derivational suffixes (e.g. *blū-a* 'blue' or *walš-itik-a* 'Welsh') take a different set of inflections from indigenous adjectives and older loans. In Piedmontese Sinti, a new xenoclitic class of adjectives has developed through the influence of the xenoclictic inflection of nouns (cf. the loan *tambl-o*, PL *tambl-i* 'dark' vs. the indigenous *kal-o*, PL *kal-e* 'black'). Borrowed adjectives, or at least adjectives borrowed from a certain source language, are uninflected in numerous dialects (e.g. German

loans in Core Sinti, Turkish loans in some dialects of the Balkans), which does not assign them a clear oikoclitic vs. xenoclitic status. Xenoclitic extension in adjectival inflection is only attested with participles in some North Central dialects. Here the xenoclitic intrusive morpheme *-on-* is extended to the oblique of indigenous participles, while their nominative inflection remains oikoclitic (e.g. the plural nominative *kerd-e* vs. oblique *kerd-on-e* 'done'). Thus, this xenoclitic extension is partial both in terms of affected lexemes and in terms of grammatical contexts.

Assuming the Early Romani distinction in the third-person singular nonperfective verb inflections between the oikoclitic *-(e)l* and the xenoclitic *-i*, a complete oikoclitic extension must have taken place in the majority of Romani dialects, which now only possess the marker *-(e)l*. Welsh and Latvian Romani have undergone a complete but variant oikoclitic extension: the oikoclitic marker is an option with all borrowed verbs, alongside the xenoclitic marker (e.g. Latvian Romani *braucin-el* or *braucin-i* 's/he drives'). According to one diachronic scenario, a type of partial oikoclitic extension has taken place in some Vlax dialects, especially Lovari, where the xenoclitic *-j* (< *-*i*) is used with the so-called contracted forms of borrowed verbs, while the oikoclitic *-el* is used with the so-called uncontracted forms of borrowed verbs (e.g. Austrian Lovari *traji-j* vs. *traji-sar-el* 's/he lives').[5] Gilan Arli, on the other hand, exhibits a partial xenoclitic extension, with the xenoclitic marker being optionally employed with polysyllabic indigenous verbs (e.g. *mudar-i* alongside *mudar-el* 's/he kills' but only *ker-el* 's/he does'); it is obligatory with borrowed verbs. Slovene Romani is the only dialect that shows an almost complete xenoclitic extension, with all verbs but a few irregular ones taking the marker *-i* (e.g. *ker-i* 's/he does').

Participles exhibit both directions of complete and partial extensions. In some Sinti and South Central dialects, borrowed verbs may form both xenoclitic and oikoclitic participles (e.g. German Sinti *xoje-men* or *xoje-do* < *xojev-* 'make angry').[6] In some Central dialects (e.g. the North Central varieties of Western Slovakia and Štítník valley, and Sípos Rumungro) and in Ukrainian Romani, the oikoclitic extension is complete and invariant in that all borrowed verbs form oikoclitic participles. Complete xenoclitic extension is typical of Finnish Romani, where all indigenous verbs form xenoclitic participles, at least variantly (e.g. *phak-ime* < *phak-* 'break'). In numerous dialects there are classes of indigenous verbs that form xenoclitic participles due to various analogies, i.e. there is a partial xenoclitic extension. In Šóka Rumungro, for example, xenoclitic participles are found with indigenous verbs whose perfective marker is homonymous to the adaptation marker *-in-* of borrowed

verbs (e.g. *čumid-* 'kiss' has the perfective stem *čumid-in-d-* and the participle *čumid-ime*, due to a partial analogy with a loan such as *ir-in-* 'turn' with the perfective stem *ir-in-d-* and the participle *ir-ime*).

Diminutives show both directions of partial extension, but only oikoclitic diminutives may completely replace xenoclitic ones. Adjectives and adverbs are open to partial extensions in both directions. We find the indigenous diminutive marker with borrowed lexemes (e.g. Central *kedvešn-or-o* < *kedvešno* 'dear', *čep-or-o* < *čepo* 'little, a little'); and borrowed diminutive markers with indigenous lexemes (e.g. Russian Romani *lač'-in'k-o* < *lač'o* 'good', Central *čul-ičk-a* < *čulo* 'little, a little'). In nouns, borrowed diminutive markers may extend to some indigenous lexemes (e.g. Russian Romani *graj-ušk-o* < *graj* 'horse', and Klenovec Rumungro *lōv-aťi* < *lōvo* 'money, coin' or Sepečides *kašt-ak-i* < *kašt* 'tree'). In western South Central dialects (Roman and western Rumungro), on the other hand, the indigenous diminutive marker has been generalised for all nouns, i.e. extended to all borrowings as well (e.g. *komuništ-ōr-o* < *komuništa* 'communist').

Female and agentive noun derivations undergo partial xenoclitic extension in numerous dialects. The extension in female nouns is usually restricted to a few lexemes (e.g. Bohemian Romani *lurd-ic-a* < *lurdo* 'soldier', Lovari and Rumungro *čōr-kiň-a* < *čōr* 'thief', Crimean Romani *amal-ink-a* < *amal* 'friend'). Although borrowed female markers appear to be productive in Sinti (e.g. *sikepaskr-ec-a* < *sikepaskro* 'teacher'), they do not completely replace indigenous female derivations. The use of borrowed agentive markers with indigenous bases may be restricted to a few items (e.g. Sinti *but-ar-i* 'worker' < *buti* 'work', Rumungro *mas-oš-i* 'meat-lover' < *mas* 'meat', Sepečides *xoxam-dži-s* 'liar' < *xoxav-* 'lie'), or it may be productive, as in the Northeastern and the North Vlax dialects (cf. Taikon Kalderaš denominal *gurumnj-är-i* 'cowboy' < *gurumni* 'cow', *luludž-ar-i* 'flower-seller' < *luludži* 'flower', *stadž-är-i* 'hat-maker' < *stadži* 'hat' etc., or deverbal *ďilaba-tor-i* 'singer' < *ďilaba-* 'sing', *phir-itor-i* 'traveller' < *phir-* 'walk, travel' etc.). Indigenous female and agentive markers do not extend to borrowed bases.

Abstract noun derivations (de-verbal and de-adjectival nominalisations) may be affected by complete extensions in both directions. The generalisation of the indigenous abstract marker *-ben ~ -pen* and their variants is found in most dialects at least in the nominative (see also Chapter 16). In some dialects, there is no trace of the Greek-derived marker *-mos*, which has been completely replaced by the indigenous marker; the latter is now used with all borrowed bases, too (e.g. Rumungro *vašal-ibe* 'ironing' < *vašal-in-* 'iron', or *mīln-ipe* 'depth' < *mīlno* 'deep'). In most North Vlax dialects, on the other hand, the

Greek-derived marker *-mos* has been generalised at the expense of the indigenous marker; it is now used with all or almost all indigenous bases, too (e.g. Taikon Kalderaš *xa-mos* 'food' < *xa-* 'eat', or *maṫ-imos* 'drunkenness' < *mato* 'drunk'). There are also dialects with extensions affecting only a few lexemes (e.g. East Slovak Romani *puča-ben* 'loan' with a borrowed base and an indigenous marker, and *kam-išag-os* 'debt' with an indigenous base and a borrowed marker).

Both directions of extension are also attested with de-substantival adjectives and de-adjectival adverbs. As for adjectives, indigenous derivational suffixes may be used with some borrowed bases in many dialects (e.g. Sinti *vul-en-o* < *vulo* 'wool', South Central *kečk-an-o* < *kečka* 'goat', Norwegian Lovari *sved-an-o* 'Swedish' < *svedo* 'Swede'). On the other hand, in Sinti, the xenoclitic suffix *-tik-* of Greek origin is employed with interal derivations of ethnic adjectives from indigenous bases (e.g. *bibol-tik-o* 'Jewish' < *biboldo* 'Jew'), although it does not extend to Early Romani derivations. As for adverbs, the xenoclitic intrusive morpheme *-on-* may extend to indigenous bases (e.g. East Slovak Romani *gul-on-es* 'sweetly' < *gulo* 'sweet', Rumungro *šūž-ōn-e* 'in a clean way' < *šūžo* 'clean'), or it may be absent with borrowed bases (e.g. Latvian Romani *darm-es* 'in vain'). Also, adverbs derived from xenoclitic adjectives in *-ik-* or *-(i)tik-* may be formed by means of the oikoclitic suffix *-es*, which had replaced the xenoclitic *-a* (e.g. Sinti *valštik-es* 'in French' < *valštiko* 'French').

23.4. Exposition

Most oikoclitic and xenoclitic markers appear to exhibit an equal or comparable degree of exposition. The only exception are the participle suffixes, where the xenoclitic value is more exposed than the oikoclitic one. The oikoclitic participle suffixes *-d-*, *-l-,* etc. also function as perfective markers (e.g. *ker-d-o* 'done' and the perfective stem *ker-d-*), while the xenoclitic suffix *-men* is restricted to the participle (e.g. *ir-imen* 'turned' vs. the perfective stem *ir-in-d-*).

23.5. Borrowing and diversity

Xenoclitic markers are always borrowed, while oikoclitic markers are never borrowed.[7] Xenoclitic markers also show greater cross-dialectal diversity than their oikoclitic counterparts. This is partly a consequence of the fact that post-

Greek xenoclitic markers are borrowed from different source languages in different dialects. Disregarding phonological developments, indigenous and Greek markers, on the other hand, retain their cross-dialectal uniformity to a great extent. Instances of a greater diversity of oikoclitic markers are rare. An example are the participle suffixes, where – disregarding dialects with complete oikoclitic or xenoclitic extensions – a single xenoclitic marker *-men* for borrowed verbs contrasts with cross-dialectally diverse ways of assigning the numerous oikoclitic markers to different classes of indigenous verbs.

Chapter 24
Criteria for asymmetry and their distribution across categories

This chapter offers a summary of asymmetries between and among categorial values classified according to the different criteria, and of the relevance of different criteria to different categories, as well as a discussion of the applicability of criteria. The order of discussion of individual categories within subsections reflects the order of previous discussion.

We distinguish four kinds of relevance of criteria to categories. First, all values of a category are consistently hierarchised by a given criterion (*complete* hierarchisation). Second, only some pairs of values of a category are consistently hierarchised by a given criterion, while other pairs of values show conflicting asymmetries or no asymmetry at all (*partial* hierarchisation). Third, some or all pairs of values of a category show conflicting asymmetries and there are no consistent asymmetries among them (*conflicting* hierarchisation). The conflict in asymmetries may be due to different asymmetries in different structures and contexts, and/or due to different asymmetries in different dialects. Finally, some criteria do not impose any salient asymmetries at all on some categories (no hierarchisation).

24.1. Complexity

The categories of number, degree, negation, cardinality, discreteness, aspect, mood, aktionsart, conditionality, transitivity, internal case (abbreviated to 'case I' in figures), indefiniteness, auxiliarity, associativity, and chronological layer are completely hierarchised by the criterion of complexity. Partial complexity asymmetries are found in the categories of tense, modality, external case (abbreviated to 'case II' in figures), case roles, localisation, ontological state, and nominal lexicality (abbreviated to 'lexicality 2' in the tables). There is no salient complexity asymmetry in the categories of evidentiality and factuality. The complexity asymmetries are charted in Table 24.1.

In tense, the remote tenses (pluperfect and imperfect) are more complex than the corresponding non-remote tenses (preterite and present), and the

future is more complex than the present, but the relation between the imperfect and the future is ambiguous: in some dialects the former is more complex than the latter, while in other dialects the latter is more complex than the former, both asymmetries occurring in a comparable number of structural types of tense systems and in a comparable number of dialects. In modality, volition is more complex than necessity, ability and inability, and necessity and ability are both more complex than inability, but the relation between necessity and ability is ambiguous. The ontological hierarchy in Table 24.1 does not involve the determiner value, which shows conflicting asymmetries with regard to the other ontological values (being the most complex value in some structures, but the least complex one in other structures); and it does not involve the place value, whose position is ambiguous.

Table 24.1. Complexity asymmetries

Category	Value asymmetry
Number	Plural > singular
Degree	Superlative > comparative > positive
	Non-positive > positive
Negation	Negative > affirmative
Cardinality	Higher > lower
Discreteness	More discrete > less discrete
Aspect	Perfective > non-perfective
Mood	Indicative > subjunctive > imperative
Aktionsart	Aktionsart modification (e.g. Iterativity) > neutral aktionsart
Conditionality	Irrealis > potential > realis
Transitivity	Transitive > intransitive
Case I	Oblique > nominative
Indefiniteness	Free-choice > negative > specific > universal (> question)
Auxiliarity	Non-auxiliary > auxiliary
Associativity	Associative > non-associative
Chronological	Xenoclitic > oikoclitic
Tense	Remote > non-remote
	Future > present
Modality	Volition > necessity, ability > inability
Case II	Genitive > other oblique cases > accusative
Case roles	Adverbial > core
Localisation	Peripheral > core
Ontological	Cause/goal > manner > thing, time, quantity > person
Lexicality 2	Noun > modifier (other than article) > definite article

The categories of person, gender, and orientation show conflicting complexity asymmetries, depending on the structure and the grammatical environment examined. In gender, the feminine tends to be more complex in nouns, but the masculine tends to be more complex in third-person pronouns and demonstratives; there are no obvious complexity asymmetries in adjectives or verbs. In person, for any pair of the three values of the category, one of the values is more complex in some structures but less complex in other structures. For example, the first person is more complex than the second person in singular personal pronouns, but the second person is more complex than the first person in plural pronouns. As for the category of orientation, there does not seem to be a simple way to generalise over various dialect-specific complexity asymmetries, which may, moreover, differ according to the structure involved.

24.2. Erosion

The categories of person and discreteness are completely hierarchised by the criterion of erosion, and there is a partial erosion asymmetry in tense, external case, localisation, and the ontological category. No salient erosion asymmetries are attested in the categories of number, degree, negation, cardinality, aspect, mood, aktionsart, evidentiality, modality, conditionality, factuality, transitivity, internal case, case roles, orientation, indefiniteness, auxiliarity, nominal lexicality, associativity, or chronological layer. For gender see below. The erosion asymmetries are shown in Table 24.2.

In the category of gender, there is some evidence that the masculine tends to undergo more erosion than the feminine in personal pronouns. This tendency, however, is probably contiguous on the original shape of the relevant stuctures, and so inconclusive for our purposes. In tense, the (remote) imperfect tends to undergo more erosion than the (non-remote) future or present-future,

Table 24.2. Erosion asymmetries

Category	Value asymmetry
Person	3 > 2 > 1
Discreteness	Less discrete > more discrete
Tense	Remote > non-remote (future or present-future)
Case II	Genitive > other oblique cases
Localisation	Inessive, contact > others
Ontological	Cause/goal > manner > determiner, thing

while there appear to be no erosion asymmetries among the other tense values. In external case, the genitive is more likely to erode than the other oblique cases. In localisation, inessive and contact prepositions appear to be more erodable than prepositions of other localisations. In the ontological category, the cause/goal value is more prone to erosion than manner, which is in turn more prone to erosion than the determiner and the thing values. There are conflicting erosion asymmetries between determiner and thing. The other ontological values do not participate in the relevant erosion development.

24.3. Differentiation

The criterion of differentiation appears to be relevant for all of our categories, except for aktionsart, case roles, and indefiniteness. The categories of negation, cardinality, discreteness, evidentiality, modality, conditionality, factuality, transitivity, internal and external case, auxiliarity, nominal lexicality, associativity, chronological layer, and also number (see below), are completely hierarchised by the criterion of differentiation. Partial differentiation asymmetries are found with the categories of person, degree, tense, mood, localisation, orientation, and ontological. The differentiation asymmetries are shown in Table 24.3.

As for the category of number, there is an overall tendency for the singular to be more differentiated than the plural (in nine out of ten contexts defined by cross-cutting category and structure), although case differentiation in nouns is exceptional in being less in the singular than in the plural. Even though there is, technically, a conflict in the differentiation asymmetries, we want to reflect this quantitative preponderance of one of the asymmetries and consider the category of number to be consistently hierarchised by the criterion of differentiation.

In the category of person, the first person is more differentiated than the second person, while there are conflicting asymmetries involving the third person. In tense, the remote tenses are less differentiated than the corresponding non-remote tenses and the future, but there is no differentiation asymmetry between the future and the (non-remote) present. In mood, the imperative is less differentiated than the indicative and the subjunctive, but there is no differentiation asymmetry between the latter two. In degree, the positive is more differentiated than the other degree values. In systems with two values, there actually is a complete differentiation asymmetry, with the positive being more differentiated than the non-positive. In systems with three degree values, however, there is no significant differentiation asymmetry between the com-

Table 24.3. Differentiation asymmetries

Category	Value asymmetry
Number	Singular > plural
Negation	Affirmative > negative
Cardinality	Lower > higher
Discreteness	More discrete > less discrete
Evidentiality	Non-evidential > evidential
Modality	Volition > necessity > ability > inability
Conditionality	Realis > potential > irrealis
Factuality	Non-factual > factual
Transitivity	Intransitive > transitive
Case I	Nominative > oblique
Case II	Genitive > accusative > other oblique cases
Auxiliarity	Auxiliary > non-auxiliary
Lexicality 2	Noun > modifier (adjective > possessive > definite article)
Associativity	Associative > non-associative
Chronological	Oikoclitic > xenoclitic
Person	1 > 2
Degree	Positive > non-positive (comparative, superlative)
Tense	Non-remote, future > remote
Mood	Indicative, subjunctive > imperative
Localisation	Core > peripheral
Orientation	Stative, directive > separative > perlative
Ontological	Person > thing > determiner > time, manner, cause/goal Quantity > place > time, manner, cause/goal

parative and the superlative, and so the asymmetry is partial. In orientation, the stative and directive are more differentiated than the separative (which is in turn more differentiated than the perlative), but there is no differentiation asymmetry between the former two values. In aktionsart, the greater differentiation of the iterative than of neutral aktionsart is marginal, both cross-dialectally and in terms of the structures involved; hence it has not been charted in Table 24.3. In the ontological category, the person value is more differentiated than thing, which is in turn more differentiated than determiner; and the quantity value is more differentiated than place. All of the previously mentioned ontological values are more differentiated than time, manner, and cause/goal, which show no differentiation asymmetry among them. There are conflicting asymmetries between person and/or thing and/or determiner on the one hand, and quantity and/or place on the other hand, which is the reason one has to construct two partial hierarchies (as shown in Table 24.3).

The categories of gender and aspect exhibit conflicting differentiation asymmetries. In gender, feminine nouns show more differentiation in number, but masculine nouns show more differentiation in class; and gender asymmetries with regard to differentiation in lexical type of adjectivals may assume both directions of prominence. In aspect, perfective forms are more differentiated in evidentiality and inflectional classification, and less differentiated in number; and the direction of aspect asymmetry in person differentiation is dialect-specific.

24.4. Extension

The categories of number, negation, discreteness, tense, aspect, transitivity, internal case, orientation, and chronological layer are completely hierarchised by the criterion of extension. Partial extension asymmetries are found with degree, mood, indefiniteness, and the ontological category. There appear to be no salient extension asymmetries in the categories of aktionsart, evidentiality, modality, conditionality, factuality, external case, case roles, localisation, auxiliarity, nominal lexicality, or associativity. The extension asymmetries are shown in Table 24.4.

Table 24.4. Extension asymmetries

Category	Value asymmetry
Number	Singular > plural
Negation	Negative > affirmative
Discreteness	More discrete > less discrete
Tense	Non-remote (present, preterite) > remote (imperfect, pluperfect) > future
	Present > future
Aspect	Non-perfective > perfective
Transitivity	Transitive > intransitive
Case I	Nominative > oblique
Orientation	Separative > stative > directive
Chronological	Oikoclitic > xenoclitic
Degree	Positive > non-positive
Mood	Indicative > subjunctive
Indefiniteness	Free-choice > specific > negative
Ontological	Place > determiner > thing > cause/goal
	Person > place

In degree, the positive may extend to the non-positive, while extensions in both directions occur between the comparative and the superlative. The category of mood involves a partial extension asymmetry; the imperative does not participate in any extension development. In indefiniteness, the free-choice value may extend to the specific value, which in turn may extend to the negative value. Both directions of extension are found between the free-choice and the universal values. In the ontological category, the place value may extend to the determiner, which may in turn extend to the thing value, which may in turn extend to the cause/goal. Although the person may extend to the place value, both directions of extension are attested between the person and the determiner, and so one must construct a separate partial hierarchy between the person and the place.

Conflicting extension asymmetries are found in the categories of person, gender, and cardinality. In person, the first person may extend to the second person as well as vice versa, depending on the structure; and the second person may extend to the third person as well as vice versa, depending on the dialect; there are no direct extensions between the first and the third persons. In gender, although in adjectivals masculine forms extend to the feminine but not vice versa, both directions of gender extension are attested in personal pronouns. In cardinality, extension in both directions, from lower numerals to higher numerals and vice versa, is attested.

24.5. Extracategorial distribution

The categories of number, person, gender, mood, and ontological value are completely hierarchised by the criterion of extracategorial distribution, and

Table 24.5. Distribution asymmetries

Category	Value asymmetry
Number	Singular > plural
Person	3 > 2 > 1
Gender	Masculine > feminine
Mood	Subjunctive > indicative > imperative
Ontological	Thing > place > manner, quantity > time > cause/goal, determiner, person
Localisation	Core > peripheral
Indefiniteness	Free-choice, negative > specific, universal

the categories of localisation and indefiniteness are partially hierarchised. There are no salient distribution asymmetries, whether complete or partial, in the categories of degree, negation, cardinality, discreteness, tense, aspect, aktionsart, evidentiality, modality, conditionality, factuality, transitivity, internal or external case, orientation, auxiliarity, nominal lexicality, associativity, or chronological layer. The distribution asymmetries are shown in Table 24.5.

As for indefiniteness, there does not seem to be any clear asymmetry between the free-choice and the negative values; or between the specific and the universal values. Conflicting distribution asymmetries are found in the category of case roles.

24.6. Exposition

The categories of number, gender, discreteness, mood, transitivity, orientation, and chronological layer are completely hierarchised by the criterion of exposition, and the categories of person, degree, tense, conditionality, indefiniteness, and the ontological category are partially hierarchised. The categories of negation, cardinality, aspect, aktionsart, evidentiality, modality, factuality, internal and external case, case roles, localisation, auxiliarity, nominal lexicality, and associativity do not exhibit any salient exposition asymmetries. The exposition asymmetries are shown in Table 24.6.

Table 24.6. Exposition asymmetries

Category	Value asymmetry
Number	Singular > plural
Gender	Masculine > feminine
Discreteness	More discrete > less discrete
Mood	Imperative > subjunctive > indicative
Transitivity	Intransitive > transitive
Orientation	Separative > directive > stative
Chronological	Xenoclitic > oikoclitic
Person	1 > 2, 3
Degree	Positive > comparative, superlative
Tense	Remote (imperfect) > future > present
Conditionality	Irrealis > realis, potential
Indefiniteness	Universal > specific, negative > free-choice
Ontological	Other > determiner, person

In the category of person, the first person is more exposed than the second and the third persons, which do not show any exposition asymmetry between them. In three-value degree, the positive is more exposed than the comparative and the superlative, but there is no exposition asymmetry between the latter two, or in the category of degree with two values.

24.7. Internal diversity

The categories of number, gender, degree, negation, discreteness, aktionsart, transitivity, internal case, orientation, ontological value, and chronological layer are completely hierarchised by the criterion of internal diversity. The category of indefiniteness shows only a partial hierarchy. The categories of person, tense, aspect, mood, evidentiality, modality, conditionality, factuality, external case, case roles, localisation, auxiliarity, nominal lexicality, and associativity do not appear to exhibit any salient asymmetries of internal diversity. The diversity asymmetries are shown in Table 24.7.

The category of cardinality shows two conflicting but internally consistent asymmetries: in cardinals the higher numerals tend to be more diverse, while in ordinals the lower numerals tend to be more diverse.

Table 24.7. Diversity asymmetries

Category	Value asymmetry
Number	Plural > singular
Gender	Masculine > feminine
Degree	Superlative > comparative > positive
	Non-positive > positive
Aktionsart	Aktionsart modification > neutral aktionsart
Transitivity	Transitive > intransitive
Discreteness	More discrete > less discrete
Case I	Nominative > oblique
Orientation	Directive > stative > separative
Negation	Affirmative > negative
Ontological	Quantity > place > determiner > person > time, thing, manner, cause/goal
Chronological	Xenoclitic > oikoclitic
Indefiniteness	Free-choice > specific, negative, universal

24.8. Borrowing

The categories of number, gender, degree, negation, discreteness, aspect, aktionsart, modality, internal case, orientation, auxiliarity, and chronological layer are completely hierarchised by the criterion of borrowing. The categories of tense, conditionality, case roles, localisation, indefiniteness, ontological value, and nominal lexicality are only partially hierarchised. There does not seem to be any salient borrowing asymmetry in the categories of mood, evidentiality, factuality, transitivity, external case, or associativity. The borrowing asymmetries are shown in Table 24.8.

In tense, the present is more likely to be borrowed than both the imperfect (the corresponding remote tense) and the future. There are no borrowing asymmetries between the imperfect and the future or between the perfective tenses (i.e. the preterite and the pluperfect). In conditionality, the realis is more prone to borrowing than the potential and irrealis, while there is no borrowing asym-

Table 24.8. Borrowing asymmetries

Category	Value asymmetry
Number	Plural > singular
Gender	Masculine > feminine
Degree	Non-positive (superlative > comparative) > positive
Negation	Affirmative > negative
Discreteness	More discrete > less discrete
Aspect	Non-perfective > perfective
Aktionsart	Aktionsart modification > neutral aktionsart
Modality	Necessity > ability > inability > volition
Case II	Nominative > oblique
Orientation	Perlative > separative > directive > stative
Auxiliarity	Non-auxiliary > auxiliary
Chronological	Xenoclitic > oikoclitic
Tense	Present > imperfect, future
Conditionality	Realis > potential, irrealis
Case roles	Adverbial > core
Localisation	Peripheral > core
Indefiniteness	Negative > specific
Ontological	Determiner, time > thing, place > person
	Determiner, time > place > quantity, cause/goal > manner
Lexicality 2	Noun > modifier (adjective > possessive, definite article)

metry between the latter two values. In indefiniteness, the negative value is more prone to borrowing than the specific value, while the position of both the universal and the free-choice values is structure-dependent. In the ontological category, two partial and partly overlapping hierarchies (as shown in Table 24.8) may be extracted from three different structure-dependent asymmetries (see Section 20.7). The determiner and the time values are more prone to borrowing than any other ontological value, while there are conflicting asymmetries between them. The place value is more likely to be borrowed than the person, quantity, cause/goal, and manner values, while there are conflicting asymmetries between place and thing. There is no borrowing asymmetry between the quantity and the cause/goal values, but they are both more prone to borrowing than the manner value. Finally, there are conflicting asymmetries between the thing and person values on the one hand, and the quantity, cause/goal, and manner values on the other hand.

The categories of person and cardinality exhibit conflicting borrowing asymmetries. In person, the second person appears to be more prone to borrowing than the first person, but the development that enables this generalisation is marginal. There are conflicting borrowing asymmetries between the third person on the one hand and the other two persons on the other hand, depending on the process: borrowing of person markers is more likely to occur in the third person, while borrowing of number markers is more likely to occur in the first and second persons. The category of cardinality shows two conflicting but internally consistent asymmetries, in that in cardinals the higher numerals are more likely to be borrowed, while in ordinals the lower numerals are more likely to be borrowed.

24.9. Criteria relevance: Summary

Table 24.9 summarises the relevance of the various criteria for asymmetry to the categories. The symbol + indicates complete hierarchisation, (+) indicates partial hierarchisation, – stands for conflicting hierarchisation, and 0 for no hierarchisation.

An impression that stands out when examining Table 24.9 is the large number of instances in which a hierarchisation is absent (indicated by 0). Before assessing the prominence of certain criteria over others, we must take into consideration some logical limits to the applicabilty of some criteria. It is possible, for instance, that some categories are inherently exempt from asymmetries that follow certain criteria. For example, extra-categorial distribution

is irrelevant for a large number of categories whose values are associated strictly with just one category, and may never be linked to others. In the absence of any asymmetry in relation to a particular criterion, the possibilities of association between criteria for a given category, and as a result the options for a clustering of categories in respect of their asymmetry behaviour, are limited.

In some cases, asymmetry hierarchies are not complete, encompassing all possible values of a category, but partial, involving only some values of the category, but not others. For example, the borrowing hierarchy 'realis > potential, irrealis' indicates that not all values can be hierarchically arranged.

Table 24.9. Summary of asymmetry criteria and their distribution across categories

Category	COM	ERO	DIF	EXT	DIS	EXP	DIV	BOR
Number	+	0	+	+	+	+	+	+
Person	–	+	(+)	–	+	(+)	0	–
Gender	–	0	–	–	+	+	+	+
Degree	+	0	(+)	(+)	0	(+)	+	+
Negation	+	0	+	+	0	0	+	+
Cardinality	+	0	+	–	0	0	–	–
Discreteness	+	+	+	+	0	+	+	+
Tense	(+)	(+)	(+)	+	0	(+)	0	(+)
Aspect	+	0	–	+	0	0	0	+
Mood	+	0	(+)	(+)	+	+	0	0
Aktionsart	+	0	0	0	0	0	+	+
Evidentiality	0	0	+	0	0	0	0	0
Modality	(+)	0	+	0	0	0	0	+
Conditionality	+	0	+	0	0	(+)	0	(+)
Factuality	0	0	+	0	0	0	0	0
Transitivity	+	0	+	+	0	+	+	0
Case I	+	0	+	+	0	0	+	+
Case II	(+)	(+)	+	0	0	0	0	0
Case roles	(+)	0	0	0	–	0	0	(+)
Localisation	(+)	(+)	(+)	0	(+)	0	0	(+)
Orientation	–	0	(+)	+	0	+	+	+
Indefiniteness	+	0	0	(+)	(+)	(+)	(+)	(+)
Ontological	(+)	(+)	(+)	(+)	+	(+)	+	(+)
Auxiliarity	+	0	+	0	0	0	0	+
Lexicality 2	(+)	0	+	0	0	0	0	(+)
Associativity	+	0	+	0	0	0	0	0
Chronological	+	0	+	+	0	+	+	+

Still, there is no conflict with the complexity hierarchy 'irrealis > potential > realis' in terms of the overall arrangement of values (the difference being in the direction of polarity, and the number of demarcations between values, but not in the overall order). Partial hierarchies can thus be integrated into the general pattern of asymmetry for individual categories: they may not confirm the main hierarchy in all its aspects, but they do not violate it, either.

With this in mind, we can turn our attention to the presence of criteria triggering asymmetry hierarchies in the various categories, as seen in Table 24.9. By far the most frequent criteria that yield complete or partial asymmtery hierarchies are Complexity and Differentiation (each relevant in 22 out of 27 categories in Table 24.9), as well as Borrowing (19 categories). We can view these therefore as the principal, most favourite strategies that are applied to structures in order to help prioritise information. Of these three strategies or criteria, Complexity and Differentation are, potentially, the more universal. Borrowing on the other hand reflects the specific way in which, in a multilingual setting, boundaries are negotiated between sets of structures within speakers' repertoires. This negotiation of boundaries is sensitive to the relations among values in categories, and so to the cognitive categorisations represented by the value opposition. The motivation for negotiating boundaries between systems (or sets within repertoires) is, however, quite distinct from the motivation to prioritise information language-internally, and we shall bear this in mind when reviewing the position of borrowing in more detail in Chapters 25 and 26. Internal diversity (relevant to 12 out of 27 categories) is an indicator of general susceptibility to change, and so it represents motivation toward renewal of some kind. Noteworthy is the almost marginal relevance of Erosion as an asymmetry criterion. This suggests either that simplification is not as prominent a process of change as one might assume, or else that, when it does occur, simplification is less subject to the hierarchical constraints that distinguish between paradigm values than are other processes of change.

Chapter 25
Patterns of asymmetry

In this chapter we consider correlations between criteria, and the behaviour of groups of categories with respect to asymmetry. We begin by examining the degree of consistency in the linear ordering of values for individual categories, in order to establish how predictable asymmetry in those categories is. We then look at recurring patterns of asymmetry among criteria and categories, and re-consider the validity of the Markedness Hypothesis in relation to the Romani data.

25.1. The consistency of value ordering within categories

25.1.1. General considerations

One of the striking findings of our survey, as outlined in Chapter 24, is that some form of asymmetry is found for most criteria, in most categories. This suggests that the arrangement of values within the categories which we examined is generally asymmetrical, implying in turn that the processes of language change that have led to the present structures of complexity, borrowing, erosion and so forth, do not affect entire paradigms in an equal fashion. Processes of change, we see, are sensitive to the hierarchical position of paradigm values, and reflect the priorities given to different kinds of information. Thus, for the bulk of the corpus of criteria and categories, patterns of the linear ordering of values can be examined, with a view toward linking these hierarchies to the conceptual motivations that condition the way information is prioritised in discourse.

Tables 25.1–25.3 present the distribution of asymmetry hierarchies for the categories examined, broken down into combinations of individual value pairs, by asymmetry criteria. The default symbol used to identify asymmetry is '>', indicating, with reference to the category value on the left, the positive presence of a property; thus, greater complexity, erosion, differentiation, diversity, etc., or more (likelihood of) borrowing or extension of the value. The symbol '><' indicates that both asymmetries are found (i.e. the value on the left is sometimes greater, and sometimes smaller, than that on the right),

and both are of comparable significance. These are the conflicting hierarchies, which will be discussed in more detail in Chapter 26. The symbol '()' indicates that some hierarchies may be observed, but they are rare or marginal in cross-dialectal comparison, and so insignificant for the sample as a whole. We use '=' to indicate that there is no asymmetry between values in respect of a criterion. The symbol '00' means that there are no relevant phenomena or developments relating to the particular criterion in a particular category, in other words, it indicates the irrelevance of a criterion as a strategy to prioritise information in a given category or sub-category, while '0' simply marks areas that are not discussed in Part Two of the book, or for which no relevant data could be identified.

The absence of relevant phenomena is apparent for the criterion of extra-categorial distribution for the categories tense, aspect, degree, and orientation (Table 25.1), as well as inflectional case (Table 25.2). As discussed in

Table 25.1. Presence of asymmetry hierarchies for a selection of categories

Category	Value order	COM	ERO	DIF	EXT	DIS	EXP	DIV	BOR
Number	SG – PL	<	0	>	>	>	>	<	<
Person	1 – 2	><	<	>	><	<	>	0	<
	1 – 3	><	<	><	<	<	>	0	<(>)
	2 – 3	><	<	><	><	<	=	0	><
Gender	M – F	><	(>)	><	>(<)	>	>	>	>
Aspect	NPFV – PFV	<	0	><	><	00	0	0	(>)
Tense	NREM – REM	<	0	>	>	00	0	0	>
	PRES – FUT	<	0	=	>	00	0	0	>
	FUT – REM	<(>)	<	>	<	00	<	0	0
Mood	IMP – SUB	<	0	<	00	<	>	0	0
	IMP – IND	<	0	<	00	<	>	0	0
	SUB – IND	<	0	<?	<	>	=	0	0
Degree	POS – COM	<	0	>	00	00	>	<	<
	POS – SUP	<	0	>	00	00	>	<	<
	COM – SUP	<	0	=(<)	>(<)	00	=	<	<
	POS – NPOS	<	0	>	>	00	00	<	<
Orientation	STA – DIR	<(>)	0	=	>	00	<	<	<
	STA – SEP	<(>)	0	>	<	00	<	>	<
	DIR – SEP	><	0	>	<	00	<	>	<
	(SEP – PER)	0	0	>	0	0	0	0	<
Aktionsart	NO – ANY	<	0	0	0	0	0	<	<
Transitivity	ITR – TR	<	0	>	<	0	>	<	0

the previous chapter, these are categories that are inherently confined to certain kinds of structures, and so their potential for extra-categorial distribution is limited. The absence of extra-categorial distribution goes along with an absence of exposition and partly also extension for the category case, where the markers for the individual values are each unique to these values, meaning that they do not derive from any synchronically co-existing structure, nor are they employed in other structures. Exposition is also irrelevant for cardinality (Table 25.1), where values are expressed through distinct lexical items, and boundaries among these items are preserved. Finally, among the ontological sub-categories (Table 25.3), we find an overwhelming tendency toward absence of extension asymmetries (as structural boundaries between the individual values are maintained), and with the exception of relations to the determiner and person values, no exposition either.

The absence of asymmetry hierarchies despite the relevance of a criterion ('=') is on the whole marginal. For example, in the category person there

Category	Value order	COM	ERO	DIF	EXT	DIS	EXP	DIV	BOR
Conditionality	REA – POT	<	0	>	0	0	=	0	>
	REA – IRR	<	0	>	0	0	<	0	>
	POT – IRR	<	0	>	0	0	<	0	=
Modality	VOL – ABI	>	0	>	0	0	0	0	<
	VOL – INA	>	0	>	0	0	0	0	<
	VOL – NEC	>	0	>	0	0	0	0	<
	ABI – INA	>	0	>	0	0	0	0	>
	ABI – NEC	<(>)	0	<	0	0	0	0	<
	INA – NEC	<	0	<	0	0	0	0	<
Factuality	NFAC – FAC			>					
Evidentiality	NEV – EV	0	0	>	0	0	0	0	0
Discreteness	LESS – MORE	<	>	<?	<	0	<	<?	<
Indefiniteness	SPEC – NEG	<	0	0	>	<	=	=	<
	SPEC – FC	<	0	0	<	<	>	<	><
	SPEC – UNIV	>	0	0	00	00	<	=	><
	NEG – FC	<	0	0	00	=	>	<	><
	NEG – UNIV	>	0	0	00	>	<	=	>
	FC – UNIV	>	0	0	><	>	<	>	><
Negation	AFF – NEG	<	0	><	<	00	0	>	>
Cardinality	L – H	<	0	>	><	0	00	<	<
(ordinal)	L – H	<	0	>	00	0	00	>	>

Table 25.2. Presence of asymmetry hierarchies for additional categories

Category	Value order	COM	ERO	DIF	EXT	DIS	EXP	DIV	BOR
Lexicality	LESS – MORE	><		><					><
Case I	NOM – OBL	<		><	>	00	00	>	>
Case II	ACC – GEN	<	<	<	00	00	00	<???	00
	ACC – OTHER	<	>?	>	00	00	00	=	00
	GEN – OTHER	>	>	>	00	00	00	>???	00
	LOC – ABL	=	=	=	><	00	00	=	=?
	LOC – SOC	=	<?	<?	00	00	00	=	>?
	ABL – SOC	=	<?	<?	>	00	00	=	>?
Case roles	CORE – ADV	<				><			<
	ATR – PRED	<		>					
	SUB/EXC – OTH	>				<?			>
Localisation	CORE – PERI	<	>	>		>			<

is exposition of both the second and third persons, as both may be extended upon, but there is no asymmetry hierarchy or greater likelihood of one of the two values, over another, to be exposed. The absence of data exemplifying a criterion stands out in connection with erosion, extension, extra-categorial distribution and partly also internal diversity for a large group of categories (Table 25.1). This does not necessarily indicate gaps in the database (although the absence of structural data will also feed into the field marked '0'); rather, it suggests that the phenomenon is in principle possible, but has not been empirically observed. For example, for the category mood, it is conceivable that erosion processes (structural reduction) could affect the values imperative, subjunctive and indicative in different ways, but in fact no phenomena of structural reduction or erosion were identified for any of these values. Similarly, for the category indefiniteness, it is possible that some values undergo greater structural reduction in some dialects, but no actual phenomena have been observed.

Finally, we turn to the absence of straightforward asymmetries, and the presence instead of conflicting hierarchies. We will discus those in more detail in Chapter 26, but here we should point out a number of criteria by which conflicting asymmetries are found in individual categories. On the whole, these are marginal, compared to the overwhelming presence of at least some consistent pattern of asymmetry with the category (albeit not always encompass-

ing all values of the category). For the criterion of complexity, we find merely conflicting hierarchies in the categories person and gender (Table 25.1), lexicality (Table 25.2), and the ontological sub-categories in relation to the value of the determiner (Table 25.3). Differentiation gives conflicting hierarchies for gender, aspect, and negation (Table 25.1), as well as for lexicality and internal case (Table 25.2). Extension only gives conflicting asymmetries for aspect and cardinality (Table 25.1), and borrowing gives no consistent hierarchies for lexicality (Table 25.2).

Table 25.3. Presence of asymmetry hierarchies in ontological sub-categories

Category	Value order	COM	ERO	DIF	EXT	DIS	EXP	DIV	BOR
Ontological	DET – PERS	><	00	<	><	=	=	>	>
	DET – THIN	><	><	<	>	<	<	>	>
	DET – PLAC	><	00	><	<	<	<	<	>
	DET – TIME	><	00	>	00	<	<	>	><
	DET – MAN	><	<	>	00	<	<	>	>
	DET – QUAN	><	00	><	00	<	<	<	>
	DET – CAUS	><	<	>	00	=	<	>	>
	PERS – THIN	<	00	>	00	<	<	>	<
	PERS – PLAC	?	00	><	>	<	<	<	<
	PERS – TIME	<	00	>	00	<	<	>	<
	PERS – MAN	<	00	>	00	<	<	>	>?
	PERS – QUAN	<	00	><	00	<	<	<	>?
	PERS – CAUS	<	00	>	00	=	<	>	><?
	THIN – PLAC	?	00	><	00	>	00	<	><
	THIN – TIME	=	00	>	00	>	00	=	<
	THIN – MAN	<	<	>	00	>	00	=	>
	THIN – QUAN	=	00	><	00	>	00	<	><?
	THIN – CAUS	<	<	>	>	>	00	=	><
	PLAC – TIME	?	00	>	00	>	00	>	<
	PLAC – MAN	?	00	>	00	>	00	>	>
	PLAC – QUAN	?	00	<	00	>	00	<	>
	PLAC – CAUS	?	00	>	00	>	00	>	>
	TIME – MAN	<	00	=	00	<	00	=	>
	TIME – QUAN	=	00	<	00	<	00	<	>
	TIME – CAUS	<	00	=	00	>	00	=	>
	MAN – QUAN	>	00	<	00	=	00	<	<
	MAN – CAUS	<	<	=	00	>	00	=	<
	QUAN – CAUS	<	00	>	00	>	00	>	=

25.1.2. Variation in linear order and polarity

The definition of a 'conflict' which we used in the previous section relates to instances where a single criterion, applied to a single category, does not render a consistent linear order of values, and so the property that is associated with that criterion cannot be consistently attributed to a particular value. Below we shall discuss another kind of conflict or inconsistency, namely that where different criteria yield different linear orders. We pointed out above that there is an overwhelming tendency towards hierarchisation and an identifiable linear order of asymmetry. We must however take into consideration that some categories have just two values, and so they will always show consistent linear ordering of these values. Thus, in the category number, the singular can be either greater than, or weaker than the plural, in manifesting a certain feature; but whatever its relationship to the plural is, in the linear ordering of the two values of the category number, singular and plural, the singular will always be immediately adjacent to the plural.

From this it is clear that the notion of *linear order* or *distinct hierarchy* does not include a specification of the *direction* of the hierarchy, or the presence of a feature in association with a particular value, but merely the ordering of the values themselves. The hierarchy 'singular > plural' in relation to the criterion of extra-categorial distribution is thus identical to the hierarchy 'plural > singular' in relation to the criterion of complexity. In order to identify the actual direction of the hierarchy, in other words, the polarity relations according to which the presence or absence of the criterion feature can be associated with one end of the scale, we attach the labels 'positive' and 'negative' with reference to a selected direction. This direction is based on the random selection of the criterion of 'complexity' as the point of reference, with the lower-ranking pole figuring on the left. Thus, for the category number, the complexity hierarchy 'singular < plural' figures as the default positive direction. The same linear order appears in negative polarity for the criterion of extra-categorial distribution (or 'NEG distribution').

Logically, a single linear ordering of values will be found (assuming asymmetry can be identified), for those categories for which only two values play a role. Table 25.4 (summarising relevant data from Tables 25.1–25.2) presents the linear order and, for the relevant criteria, polarity order for such binary categories.

Other categories that show a complexity of (often contradictory) value orderings, such as the lexicality, localisation, and the ontological categories, can be broken down into partial and consistent binary value hierarchies (see

Table 25.5), or simplified (by lumping together 'peripheral' localisation values as opposed to 'core', for example, in localisation; or 'modifiers' as opposed to 'nouns' in nominal lexicality.

Among the categories that have more than two values, some show consistent linear orderings, while others do not; the latter group shows what we call 'conflicts'. In those categories that do show consistent orderings, some may be straightforward in showing just one single hierarchy (albeit with different directions of polarity); others however may still be included among the categories with consistent orderings, even if not all values are arranged in a single hierarchy, provided that whatever partial hierarchies can be identified

Table 25.4. Linear order and relevant polarity for binary categories

CATEGORY	ORDER	CRITERIA (WITH POLARITY)
Number	SG < PL	Complexity, diversity, borrowing, NEG differentiation, NEG extension, NEG distribution, NEG exposition
Gender	M < F	NEG extension, NEG distribution, NEG exposition, NEG diversity, NEG borrowing, (NEG erosion)
Case I	NOM < OBL	Complexity, NEG extension, NEG diversity, NEG borrowing
Aspect	NON-PFV < PFV	Complexity, (NEG borrowing)
Aktionsart	NO < ANY	Complexity, diversity, borrowing
Transitivity	ITR < TR	Complexity, extension, diversity, NEG differentiation, NEG exposition
Evidentiality	NON-EVI < EVI	NEG differentiation
Discreteness	LESS < MORE	Complexity, differentiation?, extension, exposition, diversity?, borrowing; NEG erosion
Negation	AFF < NEG	Complexity, extension, NEG diversity, NEG borrowing
Lexicality 1	AUX < NAUX	Complexity, borrowing, NEG differentiation
Cardinality	L < H	complexity, diversity, borrowing, NEG differentiation
(ordinal)	L < H	Complexity, NEG diversity, NEG borrowing, NEG differentiation
Localisation	CORE < PERIPHERAL	Complexity, borrowing, NEG erosion, NEG differentiation, NEG distribution
Case role	CORE < ADVERBIAL	Complexity, borrowing

are compatible with the main linear ordering pattern. For example, for the category **person**, the full hierarchy 3 < 2 < 1 is found for NEG erosion and NEG distribution. Compatible partial hierarchies are 3, 2 < 1 for NEG borrowing, exposition; 2 < 1 for differentiation; and 3 < 1 for NEG extention. Other categories with consistent orderings (full or partial) are **internal case** (nominative < vocative < oblique for complexity; nominative < vocative; oblique for NEG extension; and nominative < oblique for NEG diversity and NEG borrowing), **external case** (accusative < other < genitive for complexity; other < genitive for erosion, diversity, differentiation), **degree** (positive < comparative < superlative for complexity, diversity, borrowing; comparative < superlative for NEG extension; and positive < non-positive for NEG differentiation and NEG exposition), **conditionality** (realis < potential < irrealis for complexity, NEG differentiation; realis < potential, irrealis for NEG borrowing; and realis, potential < irrealis for exposition) (see also Table 25.5 below).

For other categories, different asymmetry criteria give contradicting linear orders. **Modality** shows inability < ability < necessity < volition for complexity and differentiation, but volition < inability < ability < necessity for borrowing, though a uniform hierarchy can be formulated if volition is excluded. **Tense** shows the compatible hierarchies non-remote < future < remote for complexity; non-remote, future < remote for NEG differentiation; non-remote < future, remote for NEG borrowing; and future < remote for erosion and exposition; but the contradicting hierarchy non-remote < remote < future for NEG extension. In **mood**, we find the compatible hierarchies imperative < subjunctive < indicative for complexity and differentiation; subjunctive < indicative for extension; and imperative < subjunctive, indicative for exposition; but the conflicting order imperative < indicative < subjunctive for extracategorial distribution. **Orientation** shows stative < directional, separative for complexity; stative < directional < separative for exposition and borrowing; and stative, directional < separative for NEG differentiation; but directional < stative < separative for extension and NEG internal diversity. And finally, **indefiniteness** shows universal < specific < negative < free choice for complexity; universal, specific < negative, free choice for distribution; and universal < specific, negative < free choice for NEG exposition; differing in the relative positions of the values 'specific' and 'negative' from negative < specific < free-choice for extension. Compatible with both orders are specific < negative for borrowing, and other < free-choice for diversity.

Remarkably, even the conflicting cases of value hierarchies tend to show a more dominant order, composed either of full or partial hierarchies, and one that is aberrant and confined to a smaller set of criteria. This leads us to sev-

eral conclusions. First, for the categories examined, and the criteria which we applied, paradigms are usually asymmetrical. There are few and marginal instances of symmetrical structuring, relatively few cases in which a single criterion renders conflicting hierarchies, and only rather peripheral cases where asymmetry criteria are not applicable, though some criteria are more prominently represented across the various categories than others. Next, there is an overwhelming tendency for hierarchies to show consistent ordering, not only within individual criteria, but even when comparing their asymmetry behaviour by different criteria. For the sample as a whole, this finding is of course strongly biased by the presence of a large number of binary or two-valued categories, where there is only one option for linear ordering. But once partial hierarchies are considered, this tendency can be found overwhelming even in multivalued categories. Moreover, even categories that show conflicting linear ordering tend towards hierarchisation of the orders into a dominant and an aberrant or subordinated order, the first attracting the majority of criteria, the second following just a minority. The picture is complex, but on the whole relatively clear: Language structures, to judge by the Romani sample, serve to prioritise information following a rather strict prominence order of values within individual categories.

25.2. Clusters of asymmetry criteria

25.2.1. Predictions of the Markedness Hypothesis and 'well-behaved categories'

Asymmetry is prevalent, and values are arranged hierarchically, and the order of hierarchies is often predictable. But can we predict the direction of the hierarchy for different criteria? Will different strategies of prioritising information cluster in the way they target particular value poles within the hierarchy? In other words, is there a tendency for values to be treated as 'marked' or 'unmarked' in respect of a bundle of criteria? In Chapters 2 and 3 we outlined prevalent assumptions about asymmetry. We saw that some approaches take what we called the 'static' approach to markedness, assuming that there is a stable and predictable hierarchical relationship between certain pairs or even entire sets of values, at least for some prominent categories of language structure, and at least by some prominent structural criteria. This Markedness Hypothesis views unmarked values as default values, which are typically predicted to be more frequent (a criterion which we do not consider), but

also less complex, more easily eroded, more differentiated, more exposed, and more widely distributed. Different views may be encountered in relation to two other criteria which we follow, namely borrowing and internal diversity. The Markedness Hypothesis thus goes a significant step beyond our discussion so far: It postulates not only consistent linear orderings of values of individual categories following different evaluation criteria, but also predictable relations between the direction or polarity of the hierarchy, with respect to the presence/absence of the individual feature or property assigned by the criterion.

In Table 25.5 we display associations among asymmetry hierarchies, paying attention now to the order of polarity. Our notation is based on the selection of one feature as an initial point of reference. Here, we adopt the conventions of Markedness Theory (or discussions in support of the Markedness Hypothesis) and position the so-called 'unmarked' value, which lacks the feature, on the left-hand side, and the 'marked' value, which possesses the feature, on the right. As a random point of reference, we pick the most widespread criterion, namely complexity (or, in its absence, another prominent criterion or alternatively an assumption about frequency). We position the less complex value to the left, and the more complex to the right ('less complex < more complex').

Again as a random choice for procedure, we then assign a positive symbol (+) to the direction of polarity obtained for the criterion 'complexity'. Any other hierarchy that has not just a matching linear order, but also matching polarity, will also be assigned a positive symbol (+), and so will be positively associated with the (randomly chosen) base hierarchy. A hierarchy that shows the opposite direction of polarity, that is where the order of elements determined by the presence of the relevant feature is opposite to that obtained for the criterion 'complexity', will be negatively associated with complexity, and so will receive a negative symbol (–) in the notation. In reading Table 25.5, the negative symbol (–) in any one field thus simply indicates that the direction of the 'greater than' symbol that appears in the default 'Value hierarchy' slot is to be reversed for the relevant field. Fields (combinations of category with criterion) that show conflicting hierarchies, or absence of a hierarchy, are not represented in the figure.

Table 25.5 nicely illustrates the prominence of the criteria of Complexity and Differentiation in shaping asymmetries in most categories, as well as the prominence of borrowing as a process of structural change in the Romani sample. We shall return to the relations among specific criteria, and in particular to the position of borrowing, later on. First, however, we wish to examine whether there is evidence in support of the Markedness Hypothesis in our

sample. Do any categories behave in accordance with the predictions of the Markedness Hypothesis?

Table 25.5. Summary of associations between hierarchies (Base Table)

Category	Value hierarchy	COM	DIF	DIS	EXT	BOR	EXP	ERO	DIV
Number	SG < PL	+	−	−	−	+	−		+
Person	3 < 2 < 1		+	−	−	−	+	−	
Gender	M < F			−	−	−	−	(−)	−
Degree	POS < (COMP < SUP)	+	−		−	+	−		+
Negation	AFFIRM < NEG	+	−		+	−			−
Discreteness	NON-DISCR < DISCR	+	+?		+	+	+	−	+?
Cardinality	LOW < HIGH	+	−			+			+
(ordinal)	LOW < HIGH	+	−			−			−
Transitivity	ITRANS < TRANS	+	−		+		−		+
Tense (A)	NON-REM < FUT < REM	+	−			−	+	+	
Tense (B)	NON-REM < REM < FUT				−				
Aspect	NON-PERF < PERF	+			−	(−)			
Aktionsart	NEUTR < AKTIONS	+				+			+
Modality	INAB < ABI < NEC < VOL	+	+						
Mood	IMP < SUB < IND	+	+		+		+		
Conditionality	REAL < POT < IRREAL	+	−			−	+		
Animacy	INANIM < ANIM	+							
Lexicality 1	AUX < NON-AUX	+	−			+			
Lexicality 2	MODIF < NOUN	+	+						
	ART < OTHER MODIF	+	+						
	POSS < ADJ		+						
Localisation	CORE < PERIPH	+	−	−		+		−	
Orientation (A)	STA < DIR < SEP	+	−			+	+		
Orientation (B)	DIR < STA < SEP				+				−
Case I	NOM < OBL	+	−		−	−			−
Case II	ACC < OTHER < GEN	+	+					+	+
Indefiniteness	UNIV < SPEC < NEG < FC	+		+	−	(+)	−		
Ontological	PLACE < TIME		−	−		+			−
	PERS < THING	+	−	+		+	+		−
	THING < CAUSE	+	−	−	−			+	
	DET < THING < PERS		+			−			
	DET < TIME		−	+			+		−
	DET < MANNER		−	+		−	+	+	−
	DET < QUANT			+		−	+		+
	DET < PLACE			+	+	−	+		+

Table 25.6 shows the full extent to which the sample supports the presence of such 'well-behaved' categories. The top group (shaded), includes categories that follow the Markedness predictions by at least four criteria. This includes only the categories number, degree and the ontological pair thing–cause. In all of these, the more complex or 'marked' value is also the one that is less differentiated and less extended to other categories, and where applicable also less exposed and less widely distributed extra-categorially. The 'unmarked' value is less complex, but more differentiated and extended, more exposed and more widely distributed. Curiously, erosion, present for just two of the three categories, does not pattern with the prediction. For the moment, we shall leave both borrowing and internal diversity, neither of which figure prominently and unequivocally in the Markedness Hypothesis, out of the discussion.

All three hierarchies – number, degree, and thing–cause – might be associated with the (somewhat vague) notion of 'cognitive or conceptual complexity'. Certainly number and degree represent an iconic relationship, where structural complexity acts as a token of real-life material complexity, notably quantity. While in number, quantity is straightforward, in relation to degree one might posit a more abstract notion of quantity or material complexity, one which has more to do with the measure of an object against a presupposed quantity model (represented by the positive value at the extreme end of the scale). We are therefore concerned here to some extent at least with the processing of assumptions, and so with a scale of relevance of the propositional content conveyed by the construction: the positive value represents the neutral statement, the comparative and superlative represent different degrees of building on existing assumptions – the first, based on a smaller set of presupposed items, the second, based on a larger (potentially universal) set. Finally, with thing-cause we are dealing with the opposition of plain object-reference, and reference to causal chains which, similarly, are based on presuppositions. We might conclude that greater 'cognitive complexity' in this connection could be used as a cover-term for the **mental accessibility** of conceptual values representing real-life quantification (over singularity), the relationship between assumptions and complex sets (over plain assertions, or assumptions and simple sets), and causal chains (over straightforward object identification). It is here that the predictions of the Markedness Hypothesis are encountered most clearly.

A second group of categories shows greater diversity. These are categories that comply with the Markedness Hypothesis by at least three of the principal criteria (excluding borrowing and internal diversity). They include gender, internal case (nominative vs. oblique), localisation, transitivity, and

Table 25.6. 'Well-behaved' categories

		COM	DIF	EXP	EXT	DIS	ERO	BOR	DIV
Number	SG < PL	+	–	–	–	–		+	+
Degree	POS < (COMP < SUP)	+	–	–	–			+	+
Ontological	THING < CAUSE	+	–		–	–	+		
Gender	MAS < FEM			–	–	–	(–)	–	–
Case I	NOM < OBL	+	–		–			–	–
Localisation	CORE < PERIPH	+	–			–	–	+	
Transitivity	ITRANS < TRANS	+	–	–	+				+
Indefiniteness	UNIV < SPEC < NEG < FC	+		–	–	+		(+)	
Mood	IMP < SUB < IND	+	+	+	+				
Discreteness	NON-DISCR < DISCR	+	+?	+	+		–	+	+?
Ontological	DET < PLACE			+	+	+		–	+
Person	3 < 2 < 1		+	+	–	–	–	–	
Ontological	DET < TIME		–	+		+			–
Ontological	DET < MANNER		–	+		+	+	–	–
Ontological	PERS < THING	+	–	+		+		+	–
Ontological	DET < QUANT			+		+		–	+
Negation	AFFIRM < NEG	+	–		+			–	–
Conditionality	REAL < POT < IRREAL	+	–	+				–	
Cardinality	LOW < HIGH	+	–					+	+
(ordinal)	LOW < HIGH	+	–					–	–
Tense (A)	NON-REM < FUT < REM	+	–	+			+	–	
Tense (B)	NON-REM < REM < FUT				–				
Aspect	NON-PERF < PERF	+			–			(–)	
Aktionsart	NEUTR < AKTIONS	+						+	+
Modality	INAB < ABI < NEC < VOL	+	+						
Animacy	INANIM < ANIM	+							
Lexicality 1	AUX < NON-AUX	+	–					+	
Lexicality 2	MODIF < NOUN	+	+						
Lexicality 2	ART < OTHER MODIF	+	+						
Lexicality 2	POSS < ADJ		+						
Orientation (A)	STA < DIR < SEP	+	–	+				+	
Orientation (B)	DIR < STA < SEP				+				–
Case II	ACC < OTHER < GEN	+	+				+		+
Ontological	PLACE < TIME		–			–		+	–
Ontological	DET < THING < PERS		+					–	

indefiniteness. At first glance, we can make sense at least of the first four. With gender, the default status of the masculine value is familiar from citation forms, for instance, and it is not surprising to find it in the position of the traditionally 'unmarked' value. Similarly, the nominative value in internal case is the isolated citation form, as well as the one that, being the subject case, is more likely to appear in most types of predications, where the presence of an oblique entity implies the presence of at least one nominative entity, but not vice versa. In localisation, core–periphery relations may be associated with both egocentricity and (physical) accessibility, while for transitivity there is an obvious opposition of conceptual complexity, with transitive predicates implying at least two arguments, while intransitives imply just one. As for indefiniteness, the hierarchy poses, in fact, some more difficulties, and will be revisited in Chapter 26.

Noteworthy is a third group, with just two categories, mood and discreteness. Here too, three criteria cluster in accordance with the Markedness Hypothesis, showing the same value – the indicative is at the positive end of the scale for mood, the more discrete entities for discreteness – to be more differentiated, more exposed, and more extended. However, contrasting with the predictions of the Markedness Hypothesis, the same value is also more complex. Greater contextual independence is a common property of both values that appear at the positive end of the two scales – 'indicative', and 'more discrete'. Increased specificity, enabled by a combined strategy of complexity, differentiation and exposition, is a means to overcome relative weak accessibility triggered by this relatively weak contextual retrievability.

However we interpret the patterns in which criteria cluster for these categories, Table 25.6 reveals umabiguously that 'well-behaved' or even more remotely predictable asymmetry directions that follow the Markedness Hypothesis for any significant cluster of criteria account for just a fraction of the attested asymmetry hierarchies. However, we can extend our search for 'compliant' value orders by examining a somewhat narrower selection of criteria. Table 25.7 considers criteria that together may imply some degree of 'defaultness' of values: extra-categorial distribution, extension, and exposition. The emphasis is thus on category values that show unique and consistent form, and that are multifaceted, in that their forms may extend from one value to another, or even transcend category boundaries.

We find a first group of categories, where the three criteria – exposition, extension, and extra-categorial distribution – are indeed consistently associated with one another. These are, firstly, the 'well-behaved' category of number, gender, and the ontological pair determiner–place. The consistent directions

or positive association here mean that the 'default' values singular, masculine and place are all both uniquely exposed and multifaceted, that is, they serve further functions that exploit their original position in the category paradigm. Following is a second group, in which defaultness is confirmed by the association of the criteria of extension and distribution, but is not linked to exposition. Here we find person (with the third person as the multifaceted 'default', but the first person, high on the egocentricity scale, being more exposed), and once again the ontological pair thing–cause (with thing as the multifaceted 'default').

In the group that follows, the exposed value is also the one that is more likely to extend (universal indefiniteness, positive degree, discrete, indicative mood). In all these cases, these are values that represent greater semantic independence. The final group is characterised primarily by the lower ranking of the value determiner, in comparison with other ontological values, in one

Table 25.7. Categories with 'default' values

		EXP	EXT	DIS	COM	DIF	BOR	ERO	DIV
Number	SG < PL	–	–	–	+	–	+		+
Gender	M < F	–	–	–			–	(–)	–
Ontological	DET < PLACE	+	+	+			–		+
Person	3 < 2 < 1	+	–	–		+	–	–	
Ontological	THING < CAUSE		–	–	+	–		+	
Indefiniteness	UNIV < SPEC < NEG < FC	–	–	+	+		(+)		
Degree	POS < (COMP < SUP)	–	–		+	–	+		+
Discreteness	NON-DISCR < DISCR	+	+		+	+?	+	–	+?
Mood	IMP < SUB < IND	+	+		+	+			
Ontological	DET < TIME	+		+		–			–
Ontological	DET < MANNER	+		+		–	–	+	–
Ontological	DET < QUANT	+		+			–		+
Ontological	PERS < THING	+		+	+	–	+		–
Localisation	CORE < PERIPH			–	+	–	+	–	
Ontological	PLACE < TIME			–		–	+		–
Negation	AFFIRM < NEG		+		+	–	–		–
Orientation (B)	DIR < STA < SEP		+						–
Tense (B)	NON-REM < REM < FUT		–						
Aspect	NON-PERF < PERF		–		+		(–)		
Case I	NOM < OBL		–		+	–	–		–

case on the lower ranking of person compared to thing, which in turn is in line with the higher ranking of thing over cause in the first group.

We may conclude therefore that there is some reality to the notion of markedness, but only in a small number of categories, and in relation to a relatively modest set of criteria. Within this frame, typical 'unmarked' values are the singular (over plural), thing (over cause or person), positive degree (over non-positive); to a lesser extent masculine, third person, indicative, discrete, and universal indefiniteness; and even more remotely, nominative, core localisation, and intransitive. This categorisation however already presupposes considerable flexibility in the combination of criteria.

25.2.2. Correlating criteria: Types of clusters

Apart from the question of 'markedness' in the traditional sense, the corpus allows us some impression as to the interrelations between criteria, and the way categories tend to cluster for polarity (the direction of the hierarchy, or 'greater' and 'lesser' values) in respect of the various criteria. We saw above that the most prominent criteria that render asymmetry hierarchies across most of the categories considered are Complexity and Differentiation. Table 25.8 shows their respective distribution, for polarity, by category.

In accordance with the Markedness Hypothesis we would expect to find a negative correlation between Complexity and Differentiation; thus, a value that is more complex ('marked') is likely to be less differentiated. Table 25.8 confirms this for around two-thirds of the categories for which both criteria render consistent asymmetry hierarchies. However, there is a group for which a positive link between the two criteria is obtained. They do not all seem to share a common denominator. Nonetheless, independence and transparency both appear to play a role in some of the categories. Greater independence and in some cases transparency may be associated with the complex and differentiated values discrete, indicative, noun (over modifier), arguably volition (over other modals, all of which serve strictly as modals, while volition also has status as an autonomous predication). Even the genitive might be associated with this group. It is more differentiated in Romani primarily as a result of Suffixaufnahme (agreement with the head; see Chapter 16). On structural grounds, the addition of the agreement morpheme to the actual genitive case morpheme also turns the genitive into a more complex structure than competing values in the case paradigm, hence already the positive link between complexity and

Table 25.8. Links between Complexity and Differentiation

		COM	DIF	DIS	EXT	BOR	EXP	ERO	DIV
Number	SG < PL	+	−	−	−	+	−		+
Degree	POS < (COMP < SUP)	+	−		−	+	−		+
Negation	AFFIRM < NEG	+	−		+	−			−
Cardinality	LOW < HIGH	+	−			+			+
(ordinal)	LOW < HIGH	+	−			−			−
Transitivity	ITRANS < TRANS	+	−		+		−		+
Tense (A)	NON-REM < FUT < REM	+	−			−	+	+	
Conditionality	REAL < POT < IRREAL	+	−			−	+		
Localisation	CORE < PERIPH	+	−	−		+		−	
Orientation (A)	STA < DIR < SEP	+	−			+	+		
Case I	NOM < OBL	+	−		−	−			−
Lexicality 1	AUX < NON-AUX	+	−			+			
Ontological	PERS < THING	+	−	+		+	+		−
Ontological	THING < CAUSE	+	−	−	−			+	
Discreteness	NON-DISCR < DISCR	+	+?		+	+	+	−	+?
Modality 1	INAB < ABI < NEC < VOL	+	+						
Mood	IMP < SUB < IND	+	+		+		+		
Lexicality 2	MODIF < NOUN	+	+						
Lexicality 2	ART < OTHER MODIF	+	+						
Case II	ACC < OTHER < GEN	+	+					+	+
Aspect	NON-PERF < PERF	+			−	(−)			
Aktionsart	NEUTR < AKTIONS	+				+			+
Indefiniteness	UNIV < SPEC < NEG < FC	+		+	−	(+)	−		
Animacy	INANIM < ANIM	+							
Person	3 < 2 < 1		+	−	−	−	+	−	
Lexicality 2	POSS < ADJ		+						
Ontological	DET < THING < PERS		+			−			
Ontological	PLACE < TIME		−	−		+			−
Ontological	DET < TIME		−	+			+		−
Ontological	DET < MANNER		−	+		−	+	+	−
Gender	MAS < FEM			−	−	−	−	(−)	−
Tense (B)	NON-REM < REM < FUT				−				
Orientation (B)	DIR < STA < SEP				+				−
Ontological	DET < QUANT			+		−	+		+
Ontological	DET < PLACE			+	+	−	+		+

Table 25.9. Correlation of erosion and differentiation

		ERO	DIF	COM	DIS	EXT	BOR	EXP	DIV
Tense (A)	NON-REM < FUT < REM	+	−	+			−	+	
Ontological	DET < MANNER	+	−		+		−	+	−
Ontological	THING < CAUSE	+	−	+	−	−			
Person	3 < 2 < 1	−	+		−	−	−	+	
Discreteness	NON-DISCR < DISCR	−	+	+		+	+	+	+?
Case II	ACC < OTHER < GEN	+	+	+					+
Gender	MAS < FEM	(−)			−	−	−	−	−

differentiation. Semantically, however, it is this feature that enables the genitive to act also as a base for the derivation of new nouns, and so to show greater semantic independence and transparency than other non-nominative cases.

Erosion is only relevant to few categories. Interestingly, perhaps the only straightforward and consistent correlation between two criteria involves erosion, and its relation to differentiation (Table 25.9).

We encounter once again some of the category hierarchies which have figured prominently in the discussion in the previous section, surrounding the predictions of the Markedness Hypothesis. However, they do not behave in a way that conforms to the Markedness predictions when it comes to the criteria of erosion and differentiation. The prediction assumes greater frequency of unmarked values, which will tolerate or even require greater differentiation, but will also be subjected to greater erosion. Thus, the Markedness Hypothesis predicts a positive link between the two criteria. In Table 25.9, however, the link is almost consistently negative: More prone to erosion, but less differentiated, are the values cause, third person, and non-discrete, as well as remoteness, manner, and genitive.

Of these, all but the third person have so far tended to appear on the 'marked' side of the scale. Table 25.10 emphasises their correlation with complexity: all categories for which both an erosion- and a complexity-based hierarchy are applicable, with the exception of discreteness, show that the value that is more prone to erosion is also more complex (cause, genitive, remoteness).

For additional categories, as seen in Table 25.11, the value that is prone to erosion is similarly 'marked' according to the predictions of the Markedness Hypothesis, showing less extension and less differentiation (cause, non-discrete). Proneness to erosion therefore does not appear to be a property that accompanies simplex, 'default' and more multifaceted – so-called 'unmarked'

Table 25.10. Correlation of erosion and complexity

		ERO	COM	BOR	DIF	DIS	EXT	EXP	DIV
Ontological	THING < CAUSE	+	+		−	−	−		
Case II	ACC < OTHER < GEN	+	+		+				+
Tense (A)	NON-REM < FUT < REM	+	+	−	−			+	
Discreteness	NON-DISCR < DISCR	−	+	+	+?		+	+	+?
Person	3 < 2 < 1	−		−	+	−	−	+	
Ontological	DET < MANNER	+		−	−	+		+	−
Gender	MAS < FEM	(−)		−		−	−	−	−

values; if anything, the opposite is true, albeit altogether for a small number of categories, marginal within the sample.

Exposition was seen above (Table 25.7) to be one of the criteria that may cluster with extension and distribution, rendering 'default' values whose form is on the one hand stable and identifiable, as it is more likely to appear consistently and be marked out uniquely, and on the other hand multifaceted, in the sense that it is drawn upon to represent other values as well, and so it makes its way into further positions either within or even outside its home paradigm. As Table 25.12 shows, the relation of exposition to complexity, however, is not straightforward. The prediction of the Markedness Hypothesis, according to which the more exposed value is likely to be less complex, is upheld only in a smaller group, consisting of four categories, including the 'well-behaved' categories number and degree, as well as transitivity and indefiniteness. In fact, the first three of these categories, excluding indefiniteness (which however is not a counterexample, either), behave coherently also in respect of differentia-

Table 25.11. Correlation of erosion and extension

		ERO	EXT	DIF	COM	BOR	DIS	EXP	DIV
Ontological	THING < CAUSE	+	−	−	+		−		
Discreteness	NON-DISCR < DISCR	−	+	+?	+	+		+	+?
Person	3 < 2 < 1	−	−	+		−	−	+	
Gender	MAS < FEM	(−)	−			−	−	−	−
Case II	ACC < OTHER < GEN	+		+	+				+
Tense (A)	NON-REM < FUT < REM	+		−	+	−		+	
Ontological	DET < MANNER	+		−		−	+	+	−

tion. With transitivity, the value intransitive is thus more exposed, less complex, and more differentiated, and so on a par with singular and positive. For a somewhat larger group of categories, however, the exposed value is also the more complex value.

We turn finally to the relation between complexity and general susceptibility to change. We measure the latter by the polarity of hierarchies in respect

Table 25.12. Correlation of exposition and complexity

		EXP	COM	DIF	DIS	EXT	BOR	ERO	DIV
Number	SG < PL	–	+	–	–	–	+		+
Degree	POS < (COMP < SUP)	–	+	–		–	+		+
Transitivity	ITRANS < TRANS	–	+	–		+			+
Indefiniteness	UNIV < SPEC < NEG < FC	–	+		+	–	(+)		
Discreteness	NON-DISCR < DISCR	+	+	+?		+	+	–	+?
Mood	IMP < SUB < IND	+	+	+		+			
Tense (A)	NON-REM < FUT < REM	+	+	–			–	+	
Conditionality	REAL < POT < IRREAL	+	+	–			–		
Orientation (A)	STA < DIR < SEP	+	+	–			+		
Ontological	PERS < THING	+	+	–	+		+		–
Negation	AFFIRM < NEG		+	–		+	–		–
Cardinality	LOW < HIGH		+	–			+		+
(ordinal)	LOW < HIGH		+	–			–		–
Localisation	CORE < PERIPH		+	–	–		+	–	
Aspect	NON-PERF < PERF		+			–	(–)		
Aktionsart	NEUTR < AKTIONS		+				+		+
Modality 1	INAB < ABI < NEC < VOL		+	+					
Lexicality 1	AUX < NON-AUX		+	–			+		
Lexicality 2	MODIF < NOUN		+	+					
Lexicality 2	ART < OTHER MODIF		+	+					
Ontological	THING < CAUSE		+	–	–	–		+	
Case I	NOM < OBL		+	–		–	–		–
Case II	ACC < OTHER < GEN		+	+				+	+
Animacy	ANIM < INANIM		–						
Person	3 < 2 < 1	+		+	–	–	–	–	
Ontological	DET < TIME	+		–	+				–
Ontological	DET < MANNER	+		–	+		–	+	–
Ontological	DET < QUANT	+			+		–		+
Ontological	DET < PLACE	+			+	+	–		+
Gender	MAS < FEM	–			–	–	–	(–)	–

of the criteria of internal diversity and borrowing, as presented in Table 25.13. Three distinct groups can be identified. In the first, structural complexity correlates positively with diversity, and in most cases also with borrowability. Here, we encounter again the 'well-behaved' categories number and degree,

Table 25.13. Complexity and general susceptibility to change

		COM	DIV	BOR	DIF	DIS	EXT	EXP	ERO
Number	SG < PL	+	+	+	−	−	−	−	
Degree	POS < (COMP < SUP)	+	+	+	−		−	−	
Cardinality	LOW < HIGH	+	+	+	−				
Transitivity	ITRANS < TRANS	+	+		−		+	−	
Discreteness	NON-DISCR < DISCR	+	+	+	+		+	+	−
Case II	ACC < OTHER < GEN	+	+		+				+
Aktionsart	NEUTR < AKTIONS	+	+	+					
Ontological	PERS < THING	+	−	+	−	+		+	
Lexicality 1	AUX < NON-AUX	+		+	−				
Localisation	CORE < PERIPH	+		+	−	−			−
Orientation (A)	STA < DIR < SEP	+		+	−			+	
Indefiniteness	UNIV < SPEC < NEG < FC	+		(+)		+	−	−	
Negation	AFFIRM < NEG	+	−	−	−		+		
Ordinal cardinality	LOW < HIGH	+	−	−	−				
Case I	NOM < OBL	+	−	−	−		−		
Tense (A)	NON-REM < FUT < REM	+		−	−			+	+
Aspect	NON-PERF < PERF	+		(−)			−		
Conditionality	REAL < POT < IRREAL	+		−	−			+	
Modality 1	INAB < ABI < NEC < VOL	+			+				
Mood	IMP < SUB < IND	+			+		+	+	
Animacy	INANIM < ANIM	+							
Lexicality 2	MODIF < NOUN	+			+				
Lexicality 2	ART < OTHER MODIF	+			+				
Ontological	THING < CAUSE	+			−	−	−		+
Ontological	PLACE < TIME		−	+	−	−			
Person	3 < 2 < 1			−	+	−	−	+	−
Gender	MAS < FEM		−	−		−	−	−	(−)
Ontological	DET < THING < PERS			−	+				
Ontological	DET < TIME		−		−	+		+	
Ontological	DET < MANNER		−	−	−	+		+	+
Ontological	DET < QUANT		+	−		+		+	
Ontological	DET < PLACE		+	−		+	+	+	

in addition to discreteness, cardinality, aktionsart, transitivity, and external case. Some of these hierarchies may be said to reflect 'cognitive complexity'. Complexity is linked to greater susceptibility to change, if it reflects higher quantity (plural, high cardinality, non-positive degree), a larger number of participants (transitive), more explicit focus on internal structure (aktionsart), or separateness (discrete). In the case of the genitive, complexity is linked to greater differentiation as well as to erosion, a key factor in general internal diversity, while borrowing does not play a role.

A second group of categories, largely unaffected by internal diversity, shows a link between borrowability, and 'markedness' as defined by the correlation of high complexity and low differentiation. Here, the conceptual features seem more diverse, and there is no obvious common denominator. The thing value is low in saliency compared to person, non-auxiliary high in transparency and independence compared to the auxiliary, and peripheral localisation, separative orientation, and free-choice indefiniteness less accessible that their respective opposite poles. In a third group, it is the less complex and more differentiated value that is more prone to renewal, both internal, at least in some of the cases, and through borrowing. This appears to be the more frequent and the more salient (default or citation) value: the affirmative, low ordinal, nominative, as well as the non-remote, non-perfective, and real.

In conclusion, then, it is possible to identify criteria that do not show any tendencies to cluster, or which only do so selectively. Complexity and differentiation, for instance, the two most prevalent criteria, are not linked in any consistent way, defying the predictions of the Markedness Hypothesis. Nor are complexity and exposition clearly linked. On the other hand, erosion, though not particularly prominent in the sample as an asymmetry criterion, is almost always linked negatively to differentiation (affecting the categories tense, person, discreteness, and some ontological pairs). It is possible that the inflectionality of some of these categories helps preserve their structure, making them more resistant to erosion. The only counterexample to this group is the genitive, which is both prone to erosion, and the most differentiated value in the external case paradigm; its original structure tends to be more complex than that of other case markers, and erosion might be seen as a force creating a degree of phonological complexity comparable to that of the other external cases. Erosion and complexity tend to be linked in a positive way, though the evidence here is much weaker. The relevant values are cause, genitive, and remote. Here, greater erosion and greater complexity correlate with weaker accessibility; the exception, once again, is the genitive, whose erosion may

be interpreted, along the lines already suggested, as an analogy to other case markers within the paradigm.

Alongside these general observations on the relations between criteria, we can identify several clusters of categories, grouped on the basis of their displaying matching sensitivities to clusters of individual asymmetry criteria. Perhaps the most prominent of those is the group of 'well-behaved' categories. Their pole values – singular, thing, positive – are all characterised by low complexity, and high differentiation, high exposition, and greater extension and distribution potential (Table 25.6). We might regard them as simplex, unique, multifaceted, and buoyant, and so in this sense 'default' or 'unmarked' values. Conceptually, they all appear to represent entities that are, by comparison with counterpart values within their respective paradigm scales, easily accessible and cognitively simplex.

Another group of values is defined by high exposition, high extension, and high distribution, and contains the values singular, masculine, and place (Table 25.7). These are unique and buoyant, 'default' and multifaceted values, that can be interpreted as standing for salient entities. To a lesser degree, the group also includes both the third person, and the thing value, both buoyant and multifaceted through extension and distribution, which might similarly be viewed as representing conceptually salient entities; also related are the values universal indefiniteness, positive degree, discrete, and indicative mood, which are all potentially more strongly exposed and extended, and so figure as unique and buoyant. These might be interpreted conceptually as representing salient, but in comparison with their respective counterpart paradigm values also explicitly independent entities.

A further group of category values emerges that is characterised by high exposition, low complexity, and high differentiation: the singular, positive, intransitive, and universal indefiniteness (Table 25.12). These are unique, simplex, and multifaceted values and can be interpreted conceptually as salient, cognitively simplex, and independent.

Against a large and diverse group of category values that are both complex and volatile (Table 25.13), it is possible to identify a smaller group of values where absence of complexity correlates with volatility (general susceptibility to change and diversity): affirmative, low ordinal cardinality, and nominative – all representing conceptually salient and accessible entities.

In addition to the values mentioned above, there are other values that typically share low complexity and high differentiation: affirmative, low cardinality, realis, core localisation, stative orientation, nominative, auxiliary, and the ontological person value. In tendency, they stand for cognitively simple (low

quantity), salient, accessible, egocentric, and (with the exception of the auxiliary) independent conceptual entities.

A number of categories participate in more correlations than others, that is, more criteria are jointly active in defining asymmetry within the category. This means that more effort is being targeted to prioritise information on the basis of the specific set of category values. These frequently targeted, more prominent categories include number, degree, person, gender, discreteness, case, tense, transitivity, indefiniteness, and the ontological thing value. In Chapter 26 we shall discuss possible motivations for linear ordering of values within categories, for the sensitivity of categories to asymmetry criteria (or: prioritising strategies), and for the clusters of category values / criteria, as presented in this section.

25.2.5. The position of borrowing

Conscious of conflicting views of the relationship between borrowing and 'markedness', we have so far avoided drawing on borrowing as an asymmetry criterion in the discussion of criteria clustering, except as an indicator, together with the criterion of internal diversity, of general susceptibility to change, or volatility. It is now time to consider the position of borrowing. We have already seen that borrowing has a prominent presence among the categories sampled (Table 25.5 and Tables 25.1–25.3). This of course reflects the position of Romani as a language that is permanently in contact, in a sociolinguistic position where it absorbs influences, rather than donates forms (but see discussion of Romani as a donor language in Matras 2002, Ch 10; and cf. the entry 'Romani as donor language' in Bakker and Matras 2003). Given the general tendency toward borrowing, our interest in the current discussion is whether borrowing is subject to any of the strategies of prioritising information that play a role in internal structural change, in other words, whether the relations among values of categories promote or constrain borrowing in any way.

The obvious answer to the more general aspect of this question is affirmative, as can be seen from the presence of straightforward asymmetries between values in numerous categories, in respect of their borrowing behaviour. A more complex issue is the link between borrowing, and other asymmetry criteria. Is borrowing linked, in any way, to 'markedness' in the traditional notion? In Table 25.14 we compare borrowing asymmetries, with those obtained for our two most prominent asymmetry criteria in the sample, complexity and differentiation. Recall that the Markedness Hypothesis predicts a negative rela-

tion between these two criteria, with the idealised 'unmarked' value being less complex, and more differentiated. We have already seen that this relation between the criteria applies for some categories, but not for others. Table 25.14 shows that borrowing is independent of the nature of the link between complexity and differentiation.

The top group in Table 25.14 shows a link between greater susceptibility to borrowing, and 'markedness' as defined in the sense of the Markedness Hypothesis in respect of the criteria complexity and differentiation: The more complex, and less differentiated value is more likely to be borrowed. The categories affected are mainly of two kinds. The first sub-group shows quantity-oriented hierarchies (Number, Cardinality, Degree). Here, higher quantity correlates with greater borrowability. The second sub-group is somewhat less straightforward. The more borrowable and complex values peripheral localisation and separative orientation might be regarded as less accessible, or more remote from the egocentricity perspective, while the thing value is less salient compared to the person value. The lexicality hierarchy reflects the implication relationship between the borrowing of lexical items, and that of any grammatical elements, including auxiliaries, acknowledged in most formulated borrowing hierarchies.

In the lower group, the opposite relation appears. Here, it is the less complex and more differentiated value, the so-called 'unmarked' value, then, that is more likely to be borrowed. A particularly striking contrast is that between

Table 25.14. 'Markedness' and borrowing

		COM	DIF	BOR
Number	SG < PL	+	−	+
Cardinality	LOW < HIGH	+	−	+
Degree	POS < (COMP < SUP)	+	−	+
Localisation	CORE < PERIPH	+	−	+
Orientation (A)	STA < DIR < SEP	+	−	+
Ontological	PERS < THING	+	−	+
Lexicality 1	AUX < NON-AUX	+	−	+
Ordinal cardinality	LOW < HIGH	+	−	−
Tense (A)	NON-REM < FUT < REM	+	−	−
Negation	AFFIRM < NEG	+	−	−
Conditionality	REAL < POT < IRREAL	+	−	−
Case I	NOM < OBL	+	−	−
Discreteness	NON-DISCR < DISCR	+	+?	+

cardinality in cardinals and cardinality in ordinals (or ordinal cardinality). The semantic-conceptual nature of the scale, and so of the opposition between the values in each of these categories, appears to be identical (numerical quantity, in identical intervals). But the communicative function of the respective structures that embed the two categories is quite different. In cardinals, it is the less accessible values that are more borrowable (and more complex, often being compositions). In ordinals, it is the lower values, those, in fact, that are more frequent and so more salient (especially the form 'first'), that are more borrowable. This, indeed, is the theme in the other categories of this sub-group as well. The values non-remote, affirmative, realis, and nominative are all more borrowable, and can all be related to greater saliency, often greater independence, and presumably also higher frequency (at least in isolated citation forms, i.e. in 'default' positions) than their respective counterpart pole values. In Chapter 26, we shall need to revisit this observation and attempt to relate it to the conceptual motivations that trigger the employment of information-prioritising strategies.

Let us now re-examine the relationship between borrowing and 'markedness', now with the help of a different set of criteria, namely those that provide us with an indication of the degree of uniqueness, buoyancy and multifaceted character of values, or overall 'defaultness'. Table 25.15 compares borrowing with extra-categorial distribution, exposition, and extension. Here too, we find two conflicting patterns. The first shows that borrowing behaviour is negatively related to the 'defaultness' indicators, i.e. values that are less likely to show extra-categorial distribution, exposition or extension, or 'marked' values by this definition, are more likely to be borrowed. We have already encountered the borrowable values plural and peripheral localisation, which also figure in this group. The other values are all ontological values. The time value is the more cognitively complex and less accessible of the pair time–place. The other borrowable ontological values in this group are all measured against the determiner, which consistently shows high borrowability. In the second, bottom group the relation is the opposite. The more borrowable values are the so-called 'unmarked' values – those that are, by and large, more prone to extension, exposition and extra-categorial distribution: the masculine, the thing value, and the third person.

This confirms yet again that there is no obvious correlation between borrowability, and the 'markedness' status of a value – as defined in respect of the clustering of several behaviour criteria. It is interesting in this connection to note that borrowability often correlates with general susceptibility to change, as assessed by the likelihood of internal dialect diversity shown in the sample.

Table 25.15. Borrowing and 'default' status

		BOR	DIS	EXP	EXT	ERO	COM	DIF	DIV
Localisation	CORE < PERIPH	+	−			−	+	−	
Ontological	PLACE < TIME	+	−					−	−
Ontological	DET < MANNER	−	+	+		+		−	−
Ontological	DET < QUANT	−	+	+					+
Ontological	DET < PLACE	−	+	+	+				+
Number	SG < PL	+	−	−	−		+	−	+
Gender	M < F	−	−	−	−	(−)			−
Ontological	PERS < THING	+	+	+			+	−	−
Person	3 < 2 < 1	−	−	+	−	−		+	
Indefiniteness	UNIV < SPEC < NEG < FC	(+)	+	−	−		+		
Degree	POS < (COMP < SUP)	+		−	−		+	−	+
Discreteness	NON-DISCR < DISCR	+		+	+	−	+	+?	+?
Orientation (A)	STA < DIR < SEP	+		+			+	−	
Tense (A)	NON-REM < FUT < REM	−		+		+	+	−	
Conditionality	REAL < POT < IRREAL	−		+			+	−	
Cardinality	LOW < HIGH	+					+	−	+
Aktionsart	NEUTR < AKTIONS	+					+	(+)	+
Lexicality 1	AUX < NON-AUX	+					+	−	
Negation	AFFIRM < NEG	−			+		+	−	−
Ordinal cardinality	LOW < HIGH	−					+	−	−
Aspect	NON-PERF < PERF	(−)			−		+		
Case I	NOM < OBL	−			−		+	−	−
Ontological	DET < THING < PERS	−						+	

Table 25.16 shows that for most grammatical categories for which both criteria are relevant, the more borrowable value is also the more diverse internally.

A sub-group of categories that fall under this grouping are conceptually connected to the expression of quantity: number, degree, cardinality, and ordinal cardinality. Note once again the difference between cardinality, where the higher values are more borrowable and more diverse, and ordinal cardinality, where it is the lower value that is more borrowable and internally diverse. In a further sub-group, we find borrowing and diversity as features of the 'default', salient or independent value: masculine, nominative, affirmative, neutral aktionsart. Strikingly, the bottom group of categories, in which there is a negative correlation of borrowing and internal diversity, consists entirely of ontological value pairs. Here, the conceptual oppositions must be dealt with

individually, in relation to each of the pairs. The more borrowable values are the less-salient thing (over person), the more complex time (over place), and the more deictic and less semantically specified determiner (over quantity and place; and also over manner). In other words, in the ontological domain, saliency, simplicity, and semantic specificity may constrain borrowing.

All this continues to point to individual motivations for borrowing that are not clearly or consistently linked to other, language-internal strategies of prioritising information through structural composition of forms representing particular poles on a scale of values. Nonetheless, borrowing patterns, as we

Table 25.16. Borrowing and internal diversity

		BOR	DIV
Number	SG < PL	+	+
Degree	POS < (COMP < SUP)	+	+
Cardinality	LOW < HIGH	+	+
Ordinal cardinality	LOW < HIGH	−	−
Gender	MAS < FEM	−	−
Case I	NOM < OBL	−	−
Negation	AFFIRM < NEG	−	−
Aktionsart	NEUTR < AKTIONS	+	+
Discreteness	NON-DISCR < DISCR	+	+?
Ontological	DET < MANNER	−	−
Ontological	PERS < THING	+	−
Ontological	PLACE < TIME	+	−
Ontological	DET < QUANT	−	+
Ontological	DET < PLACE	−	+
Lexicality 1	AUX < NON-AUX	+	
Localisation	CORE < PERIPH	+	
Orientation (A)	STA < DIR < SEP	+	
Indefiniteness	UNIV < SPEC < NEG < FC	(+)	
Tense (A)	NON-REM < FUT < REM	−	
Aspect	NON-PERF < PERF	(−)	
Conditionality	REAL < POT < IRREAL	−	
Person	3 < 2 < 1	−	
Ontological	DET < THING < PERS	−	
Transitivity	ITRANS < TRANS		+
Case II	ACC < OTHER < GEN		+
Orientation (B)	DIR < STA < SEP		−
Ontological	DET < TIME		−

have already established, are also asymmetrical. Let us then at this point consider a summary of the those values that figure at the top of the respective category value hierarchies for borrowing (see also Table 24.8).

Borrowability may be positively related to a series of different conceptual features:

- Higher quantity: plural, non-positive degree, high cardinality
- Lower accessibility and greater semantic complexity: discreteness, peripheral localisation, free-choice indefiniteness, time (over place)
- Greater independence or contrast: discreteness, non-positive degree, separative orientation
- Transparency: non-auxiliary lexicality
- 'Defaultness', saliency, frequency: low ordinal cardinality, masculine gender,

Table 25.17. Polarity of borrowing

		BOR
Number	SG < PL	+
Degree	POS < (COMP < SUP)	+
Cardinality	LOW < HIGH	+
Discreteness	NON-DISCR < DISCR	+
Aktionsart	NEUTR < AKTIONS	+
Localisation	CORE < PERIPH	+
Orientation (A)	STA < DIR < SEP	+
Indefiniteness	UNIV < SPEC < NEG < FC	(+)
Lexicality 1	AUX < NON-AUX	+
Ontological	PERS < THING	+
Ontological	PLACE < TIME	+
Ordinal cardinality	LOW < HIGH	−
Gender	MAS < FEM	−
Case I	NOM < OBL	−
Person	3 < 2 < 1	−
Negation	AFFIRM < NEG	−
Tense (A)	NON-REM < FUT < REM	−
Aspect	NON-PFV < PFV	(−)
Conditionality	REAL < POT < IRREAL	−
Ontological	DET < MANNER	−
Ontological	DET < QUANT	−
Ontological	DET < PLACE	−
Ontological	DET < THING < PERS	−

nominative case, third person, affirmative, non-remote tense, non-perfective aspect, realis conditionality, determiner

There appears to be a split between the first three clusters of properties, which tend to pertain to a demarcation of intensity of reference of some kind, with greater borrowability correlating with a need to overcome conceptual boundaries; and the final clusters of features, which imply almost the opposite. We shall return to a discussion of this state-of-affairs in connection with motivations for asymmetry in Chapter 26.

Chapter 26
Conceptual motivations for asymmetry

Our point of departure in this investigation (Chapters 2 and 3) was the prevailing view in markedness theory and beyond, that categorisations in language are not accidental, but reflect the cognitive background of the conceptualisation and categorisation of reality. To this notion, we have aimed at adding a more dynamic view of asymmetry, regarding it as the outcome of strategies that are applied to individual structures, rather than to the presence of a category as a whole, and within individual categories and groups of categories, rather than globally to all types of semantic categories. In Chapter 25 we saw evidence that the interaction of criteria of asymmetry – in our view, the various strategies of prioritising information – is not global; rather, it is possible to identify various clusters of criteria and categories. Categories, in other words, react differently to combinations of asymmetry criteria. Sometimes, asymmetry hierarchies are not even consistent for the same category and for the same criterion. In the present chapter we try and make sense of the patterns found, relating them to conceptual motivations to prioritise information.

26.1. Iconic motivations for linear ordering

It is generally accepted that iconicity is a powerful motor driving the shaping and re-shaping of linguistic structures. In our sample, iconicity can be broken down to a set of several principles: quantity, immediacy, prominence, truth and simplicity, and transparency. The distribution of categories on this basis is illustrated in Table 26.1.

26.1.1. Quantity

The *iconicity of quantity* is the use of a greater quantity of linguistic structures to represent a greater number of demarcated entities, or an otherwise measurable quantity. The most straightforward strategy to represent quantity is complexity (Table 26.1). An iconic motivation relating to quantity may be assigned on the basis of the complexity criterion to the value hierarchies of

Table 26.1. Distribution of categories by iconicity principles

Category	Value hierarchy	COM	DIF	EXP	EXT	DIS	ERO	BOR	DIV
Quantity									
Number	SG < PL	+	–	–	–	–	+	+	+
Degree	POS < (COMP < SUP)	+	–	–	–			+	+
Cardinality	LOW < HIGH	+	–					+	+
Ordinality	LOW < HIGH	+	–					–	–
Aktionsart	NEUTR < AKTIONS	+	(+)					+	+
Immediacy									
Tense (1)	NON-REM < REM	+	–	+			+	–	
Tense (2)	NON-REM < FUT	+	–	+	–		+	–	
Orientation (1)	STA < DIR < SEP	+	–	+				+	
Aspect	NON-PERF < PERF	+			–			(–)	
Localisation	CORE < PERIPH	+	–			–		+	
Indefiniteness	UNIV < SPEC < NEG < FC	+		–	–	+		(+)	
Person	3 < 2 < 1		+	+	–	–	–	–	
Discreteness	NON-DISCR < DISCR	+	+?	+	+		–	+	+?
Prominence									
Case I	NOM < OBL	+	–		–			–	–
Ontological	PERS < THING	+	–	+		+		+	–
Transitivity	ITRANS < TRANS	+	–	–					+
Animacy	INANIM < ANIM	+							
Gender	MAS < FEM			–	–	–	(–)	–	–
Person	3 < 2 < 1		+	+	–	–	–	–	
Truth and simplicity									
Negation	AFFIRM < NEG	+	–		+			–	–
Conditionality	REAL < POT < IRREAL	+	–	+				–	
Transitivity	ITRANS < TRANS	+	–	–					+
Orientation (1)	STA < DIR < SEP	+	–	+				+	
Ontological	PERS < TIME < MAN < CAUSE	+	–		–	–	+		
Ontological	PLACE < TIME		–			–		+	–
Transparency									
Discreteness	NON-DISCR < DISCR	+	+?	+	+		–	+	+?
Mood	IMP < SUB < IND	+	+	+	+				
Modality	INAB < ABI < NEC < VOL	+	+						
Lexicality 2	ART < MODIF < NOUN	+	+					+	
Auxiliarity	AUX < NON-AUX	+	–					+	

number (plural > singular), cardinality and ordinality (higher > lower), as well as degree (superlative > comparative > positive) (cf. Table 24.1). To some extent, aktionsart modification might also be regarded as an expression of quantity, as in some dialects at least it relates to iterativity or intensity, and so repetition of an action. Two of the categories, number and degree, constitute two out of three 'well-behaved' categories that are maximally consistent with the Markedness Hypothesis. There is just a small number of quantity-related hierarchies, and just a small number of well-behaved hierarchies. The link between the two therefore stands out, and tells us something about the role of quantity in biasing anticipation about linguistic asymmetry: quantity hierarchies are closest to idealised hierarchies, possibly because idealised hierarchies depart from the kind of correlations between criteria that are found in quantity-iconicity hierarchies.

26.1.2. Immediacy

The second type of iconicity hierarchy is what we call the *iconicity of immediacy*. There are two dimensions that play a role here. The first is the measuring of immediacy in relation to the deictic centre, which overlaps with the speaker's position. Egocentricity motivates the assessment of entities and events in deictic terms. In prioritising information, preference is given to entities that appear more reachable from the speaker's point of view. The second dimension has been referred to above (Chapter 3) as accessiblity, and pertains to the immediacy of information from the overall contextual point of view. The measure of accessibility is therefore in the immediacy of information to both speaker and hearer, based on general knowledge as well as the active contextual knowledge of the ongoing interaction.

What do we mean by giving 'preference' to immediate information? We can begin again with the criterion of complexity. Iconicity of immediacy prescribes that less linguistic material is necessary to portray immediate information, and more to portray less immediate information. This is reflected in the lower complexity of those values that convey deictic and contextual overlap with time and place in the categories tense (for complexity: remote > non-remote, and future > present), aspect (perfective > non-perfective), and localisation (peripheral > core), as well as orientation (separative > directive > stative). In the latter, the hierarchy reflects the deictic position of the stative value, followed by the intentionality and hence lower immediacy of the directive, followed then by the weaker verifiability of origin and so the

lower-ranking position of the separative. Immediacy hierarchies may also be hearer-directed. In indefiniteness (free-choice > negative > specific > universal > question), the most complex value represents the greatest difficulty in establishing a clear and positive idea of the referent, while the least complex is the one best known to the hearer, though not necessarily to the speaker. In discreteness (more discrete > less discrete), the more complex value may be associated with contrast, with general unexpectedness (such as purpose clauses involving low subject control), or with specificity, indicating greater speaker-sided effort to win over the hearer's point of view.

The *immediacy* hierarchies form less of a tight cluster, once we examine other criteria. Nonetheless, some patterns do emerge. Thus, there is an overwhelming tendency for 'immediate' values to be more differentiated (non-remote tense, core localisation, stative orientation), non-discrete entitities being the only exception. Immediate values are also more prone to extension (non-remote tense, non-perfective aspect, universal indefinite), non-discrete again being an exception (due primarily to the extension along the deictic–anaphoric–determiner chain). The less-immediate values remote, future, and separative are all more exposed. In the category Person, there is a clash between the first person, higher on the egocentricity scale (more differentiated and more exposed), and third person, which is high in continuity in discourse, and so high in accessibility (more extended, more widely distributed, and more eroded). Significantly, there is no consistent Person hierarchy for complexity.

26.1.3. Prominence

Our next iconicity principle is the *iconicity of prominence*. It assumes that certain entities are inherently more salient topics of conversation. On the complexity scale, more salient entities appear less complex: Relevant categories are case (for complexity: oblique > nominative) and the ontological value of person (over thing). In transitivity, the less complex and more differentiated value is the intransitive, in part due to the fact that here the actor is often the undergoer. This reflects a tendency to view the outcome of actions of motions or change of state from the specific perspective of the undergoer, while transitive actions are viewed wholesale from the perspective of the agent. The prominent values – nominative, person, intransitive – are also the more differentiated.

For animacy, only complexity is relevant, and here the hierarchy seems at first glance to contradict the pattern: animate entities, the more discourse-

prominent, are also more complex. This however is due to the specific use of complexity in marking out the oblique role of the animate entity, as the role that contradicts expectations. In this respect, the hierarchy of complexity in animacy is, strictly speaking, a hierarchy of expected roles: the more complex value represents the more unexpected role (animate oblique), while the less complex value represents the expected and so more immediately accessible role (inanimate oblique).

Two additional categories figure marginally in the prominence domain. The first is gender. Masculine, more prominent in discourse, appears as 'default' in the sense that it is more consistently unique (more exposed), more buoyant at the paradigm level (more prone to extension), and has a potentially multi-faceted character (more prone to extra-categorial distribution); it is also more eroded, and more volatile (prone to borrowing and internal diversity). But significantly, gender remains unaffected by the two most frequent strategies of prioritisation, complexity and differentiation. This may alert us to the rather vague metaphorical, after all, relationship between gender and any form of 'defaultness', in contrast with, for instance, the much more plastic relationship between cardinality or number and quantity, or between core localisation or non-remote tense and immediacy. Next, the competition between values of the category person might be attributed in part to the discourse prominence of the third person, strengthening its position – much like the masculine value – especially as a 'default' value, prone to extension, distribution, erosion, and general renewal.

26.1.4. Truth and simplicity

Our next principle is what we call the *iconicity of truth and simplicity*. Conversational maxims, especially the maxims of quality and of quantity, create an expectation that default propositions are true, and concise. The criterion of complexity allows us to identify strategies of marking out the opposite poles to this default content. We find complexity in negation (negative > affirmative) and conditionality (irrealis > potential > realis), both following from the expectation that the default value should be true, affirmative and unconditional. In negation, there is also an iconic reflection of the fact that negation presupposes the potential for affirmation. Complexity is also an iconic token of greater conceptual complexity, with which we mean the involvement of a greater number of entities, propositional units, or dimensions in the conceptualisation triggered by the structure. This can be seen in transitivity (for

complexity: transitive > intransitive), where the transitive value involves a larger number of arguments, in orientation (separative > directive > stative), where the separative and directive involve a greater number of reference points than the stative, and in the ontological hierarchy (cause/goal > manner > thing, time, quantity > person), where the more complex values are those that involve linking propositions (cause, goal, and a tighter link in manner), or linking entities with external points of reference (time, quantity).

26.1.5. Transparency

Finally, we arrive at the principle of *iconicity of transparency*. This group is unique in drawing on both greater complexity and greater differentiation to mark out values that are more transparent in terms of their referential autonomy and consistency, the principle being that with weaker transparency of the structure, there is more reliance on the contextual environment, and hence less need for complex identification structures, or for differentiating features. Discreteness figures here, as discrete entities have more independent referential status. Discrete entities are consistently complex, differentiated, exposed, and extended. Similarly, mood shows a transparency hierarchy for the same four criteria (indicative > subjunctive > imperative), which mirrors the independence of the indicative value, the contextual dependency of the subjunctive, and the illocutionary or speech act dependency of the imperative. Transparency is also a feature of lexical categories. In modality (volition > necessity, ability > inability), volition exhibits greatest contextual independence (often independent of a modal function), while necessity and ability represent relative autonomy of action, against inability. Nouns figure higher than lexical modifiers, which in turn are higher on the hierarchy than articles, and non-auxiliaries are higher than auxiliaries.

26.2. Global and local motivations

Above we saw that several principles of iconicity may be held responsible for the behaviour of a large number of categories. More specifically, these principles help explain (a) the linear ordering of values; (b) the polarity direction in respect of the (most prominent) criteria or prioritisation strategies of complexity and differentiation; and (c) quite often, also the polarity direction with respect to additional criteria. The latter, the clustering of categories in respect

of additional criteria, is clearly the most problematic and irregular issue. Our dynamic approach to asymmetry predicts that prioritisation strategies will be set not just at the global level of iconic representation of categories, but also at a local level. Let us now examine some of the relevant issues.

26.2.1. Exposition

Exposition is one of the strategies that emerges from Table 26.1 as less regular, and less predictable. In fact, it patterns fairly regularly in the groups marked out for iconicity of quantity, immediacy, and transparency. In quantity, the more exposed value represents lower quantity, a token of the stability of the more simple and more easily perceivable value. This is in line with the greater extension and extra-categorial distribution of the lower value, as well as with the greater volatility of the higher value, and its greater proneness to borrowing and internal diversity. Exposition is also a strategy to mark out values that are high on the transparency scale, correlating with complexity, differentiation, and extension. Specifically, the contextual independence associated with the values discrete and indicative carries with it greater exposition.

With immediacy, exposition is positively related to complexity. Here, it figures as a token of the demarcation boundary of the less immediate value. The exception is indefiniteness, where the universal value is the most accessible/immediate, encompassing all possible entities in the respective ontological domain (eg. 'everywhere', 'always'). Although we group the universal value as immediate due to its accessibility, the concept of immediacy differs here in that entitites encompassed by the universal values are accessible by virtue of being conceptualised wholesale. The contextual overlap therefore is derived from a global generalisation, rather than from the immediate deictic-contextual centre of mental orientation. The exposition of the universal may be interpreted as a token of this globality.

At first glance it appears that a more diversified use of exposition is found with prominence, and the truth and simplicity scales. In fact, exposition is a consistent strategy of marking out lower prominence, and lower truth and simplicity. The one exception is transitivity, which figures in both domains. Here, the exposition of the intransitive is a means of identifying the various concrete semantic effects of an action or event on the actor or undergoer. As a result, unique constructions stand out as tokens of the intransitive. In the 'non-aligned' category of person, it appears that egocentricity and so immediacy favour exposition.

26.2.2. Extension, distribution, and erosion

We have been treating extension as buoyancy of a value at the paradigm level, allowing it to assume further roles, and extra-categorial distribution as a multifaceted potential of a value, allowing it to serve as base for functions outside its own category. Extension is straightforward in the groups quantity, immediacy, prominence, and transparency: although extension potential is missing altogether in many categories, it is always the lower quantity, more immediate, more prominent and more independent value that is extended. This general pattern is disturbed only by the extension of the negative value in negation (truth and simplicity). The extension of the negative, which is confined to a small number of dialect cases, results from the need to maintain a greater complexity of the negative value, coupled with the need to renew the affirmative: a more transparent negator is added to a negative modal with cumulative marking of negation and modality, and the inherent negative value of the predicate is hyperanalysed as a property of the negator alone (e.g. Sinti *naj* 'cannot' > *naj gar* 'cannot' > *naj* 'can').

Distribution is an even less prominent feature. In Table 25.7 in the previous chapter we saw that many of the distribution hierarchies pertain to the weaker distribution of the ontological determiner value, compared to other ontological values. These hierarchies are now excluded from Table 26.1. Transparency scales are, interestingly, not at all affected by distribution. In the ontological domain, simplicity favours extra-categorial distribution, though Thing outranks Person, presumably due to its lack of semantic specificity and so its greater adaptation to other category environments. In the indefiniteness category, the greater distribution of the free-choice value, which again lacks specificity, might indicate a similar motivation. Elsewhere, greater distribution is associated with low quantity (singular, in number), and immediacy (core, in localisation). In the 'non-aligned' categories person and gender, discourse prominence favours both extension and extra-categorial distribution (of the third person and masculine, respectively).

Erosion is a marginal strategy in the sample. It tends to affect higher quantity (plural), less immediate (non-remote), non-discrete, and ontologically simplex values, on the one hand. On the other hand, in the 'non-aligned' categories gender and person, the eroded values are those that have greater discourse prominence, and possibly greater frequency – those that are also more prone to extension and distribution: third person and masculine.

26.2.3. Borrowing

In Table 25.16 we saw a summary of the polarity directions for borrowing. There were two overall themes. On the one hand, more borrowable values can represent values that involve greater intensity of processing, as they are related to overcoming conceptual boundaries of some kind. The quantity-related values plural, non-positive degree, high cardinality, and aktionsart modification such as habituality are all more borrowable (though ordinal cardinals show the opposite tendency to cardinals). In the immediacy domain, the less reachable periphery and the more contrastive separative are borrowable, as are the most discrete entities and the more complex ontological time value. On the other hand, many default values are more borrowable. In quantity, lower ordinals; in immediacy, non-remote and non-prefective; in prominence, nominative, masculine, and third person; in truth and simplicity, affirmative and realis; and in transparency, greater lexicality. What might motivate borrowing?

In order to answer the question, we must return to the actual distribution of borrowed values among structures. Quite clearly, borrowing is less connected to iconic motivations, and so it is found to have the weakest correlations with the iconicity principles that work rather well for the other strategies. This reveals something about the process of borrowing itself. It is not a strategy for prioritising information at the same level as the other asymmetry criteria; rather, it is a strategy of supporting the bilingual speaker in successfully managing language choices. Borrowing, in effect, reduces the need for choices to be made among alternate systems, and so it increases communicative efficiency in bilingual situations, without compromising wholesale the separation of languages, and so the flagging of a separate identity via language. The question that is of interest to our immediate discussion is which conceptual motivations trigger the licence to compromise bilingual choices around individual structures, for borrowing too is arranged asymmetrically within categories, as we can clearly see from the presence of hierarchies of borrowing in the sample. In addressing this specific question, we must turn our attention to those constellations in communication where it is most advantageous for speakers not to be confronted with choices between systems, or in other words, where repertoire reduction is most economical.

The first factor motivating borrowing is the potential of interactional tension around elements that demand increased intervention with hearer-sided processing. These are instances where the propositional content runs contrary to expectations, and so greater mental effort is needed in order to monitor the

hearer's reaction and, if necessary, support the speaker's position. Such occurrences involve an explicit processing of hearer-sided expectations. We find this in the category degree, where non-positive values inherently involve a statement about shared knowledge concerning a prototype, and the superlative even draws a contrast between a single entity, and a set of known, potential entities. We find, of course, a similar kind of contrast in the various concrete representations of discreteness, such as contrast, change, and unexpectedness. The affirmative is high on the borrowing scale due to the borrowability of the affirmative particle itself, which acts like an emphatic discourse marker, while the negative particle remains associated with the verbal negator. The common denominator here is relevance: the higher the position of a structure on the scale of relevance (i.e. the explicit processing of hearer expectations), the greater its likelihood to be borrowed.

Borrowing, as we said above, is a strategy to help prioritise information specifically in the language contact situation, where the speaker has a social motivation to maintain a separation of repertoires, while on the other hand the effectiveness of handling communication may motivate a fusion of the systems for some structures. High conversational tension around structures that convey relevance may support the motivation to reduce the need to choose and so to discriminate between portions of the bilingual repertoire, and so to increase fusion, or non-separation of the two systems. The social constraints in such a situation will of course dictate that if two structures merge, it will be that from the dominant, prestigious language that is more likely to take over, since the minority will tolerate and accept forms from the majority language, but not vice versa. But while the choice of one form or another is determined by the social status of the languages, the motivation toward merger of subcomponents in the first place is a communicative one.

Weaker accessibility can be intergated into this motivation: the more difficult it is for the speaker to identify an object or direction, the greater the mental effort that is needed to keep the hearer's attention, the greater the need to simplify the bilingual repertoire by allowing sub-systems or components to merge (by adopting one form or structure over another). The borrowability of peripheral over core localisations, and of high cardinality over low cardinality, may be explained in this way. With cardinality, salient points of reference, which are more easily accessible to the speaker, such as 'ten' and 'hundred', are usually exempted from borrowing. This seemingly formal exception to the hierarchy, actually supports our explanation of the hierarchy. In the case of number, it is the attention given to the singular, as seen in its default value and

general stability, which makes the plural more prone to borrowing as well as general renewal.

It is clear that a different kind of motivation is responsible for the borrowability of lexical items, one which has to do with negotiating specific cultural contexts. Elements that are high on the lexicality scale are therefore more borrowable. We are left however with those values that are more borrowable, but display lower quantity (lower ordinality), greater immediacy (non-remote, non-perfective), and greater prominence (nominative, third person) and truth (realis). How can they be reconciled with the relevance-related trigger for borrowing? In most of these cases we are talking about grammatical structures that accompany borrowed lexical items, nouns or verbs (nominative, non-remote, non-perfective, third person). Their borrowability can be assumed to reflect their default status and frequency in discourse in the source language. In the case of the realis, the borrowed structure in question is a conjunction, and here too frequency may hold the key, especially in view of the low frequency of irrealis in spoken discourse. As for ordinality, the hierarchical polarity that favours lower ordinals in borrowing is in fact a reflection of the borrowability of the ordinal 'first'. What stands out is the tendency to mark out – or prioritise – this word, by treating it as separate from the composition of number + ordinal morpheme. This tendency is underlined by the high internal diversity of the ordinal 'first', and is mirrored in many other languages by the tendency of 'first' toward suppletion or other irregularity. Quite unlike the cases of fusion of repertoires or relevance-related borrowing, in the case of ordinality borrowing is simply an additional prioritising strategy that is available in a language contact situation.

26.3. Conflicting hierarchies and conflict resolution

In the previous sections we argued for some degree of internal coherence, and so for some predictability, of the linear hierarchies of individual categories and their polarity in respect of specific strategies (i.e. asymmetry criteria). This argument is based on an interpretation of the conceptual significance of individual values within a category, and of their relation with one another, as well as on a review of the functions of structures within which categories are represented. We have pointed out that Markedness is not always a consistent and pre-determined setup, and criteria may cluster differently in different categories, and may influence the direction of polarity in different ways. We have

tried above to make sense of some of those clusters, thereby restoring some confidence in the overall notion of some systematicity behind the application of strategies to prioritise information; this is our 'dynamic' model of asymmetry in language.

But we have so far concentrated our attention on those instances in which there is a clear linear hierarchy among values of a given category, for a given criterion. In the present section we approach the remaining 'fields' in the asymmetry table: those, where linear order is inconsistent. These instances deserve our attention, since the notion of dynamic asymmetry hinges on local motivations and local strategies – 'local' being a specific combination of factors involving the value, the structure, and the strategy (see Chapter 3). We call them 'conflicting hierarchies', for they do indeed manifest asymmetry, but the asymmetry pattern is inconsistent.

Some conflicting hierarchies are considered below as 'resolvable', in the sense that they can be explained by a local interplay of factors as exceptional to the picture as a whole; the remaining, default, or 'expected' hierarchy may feed into non-conflicting hierarchies established on the basis of other criteria. We are of course aware of the potential methodological, heuristic problem: We know what we expect, i.e. we know which of the conflicting hierarchies should be the feeding ones, and so we concentrate on providing local explanations for the other ('misbehaved') hierarchies. Nonetheless, we believe that this approach to conflict resolution in hierarchies is still consistent with our overall assumption that patterns are set in response to local motivations, and so if a local motivation for exceptional arrangements can be identified, it can be drawn upon in an explanatory analysis.

26.3.1. Conflict domains

We find different types of conflict in the sample. First, there are *dialect* conflicts: conflicts that arise due to contradictions between individual dialects in the linear order of category values for individual criteria. These are found for numerous categories; we have identified them for the categories person, gender, degree, tense, aspect, modality, orientation, indefiniteness, ontological value, cardinality – but there seems to be no constraint in principle on their categorial distribution. Dialect conflicts are difficult to resolve, since we cannot of course attribute different conceptual motivations to different communities of speakers. We must therefore accept that wherever the sample yields conflicting hierarchies due to different behaviour among dialects, we can either

rely on overall trends and tendencies (as we do in Part Two of this book), or, when no such general trends can be observed, we must exclude the individual cases from the discussion.

Interestingly, dialect conflicts appear for all asymmetry criteria with the exception of borrowing. This is a rather unexpected and so noteworthy outcome; one might have thought that if anything, borrowing behaviour might be significantly influenced by the structures of the source or donor language, and different among the dialects, which show a diverse range of contact languages. The fact that this is not the case strengthens our argument in favour of a cognitive (albeit discourse-managerial, rather than conceptual) motivation for borrowing, where structural solutions are found wherever needed, acting in the service, as it were, of pressure toward a fusion of sets or sub-sets of functions.

A structural conflict domain is that of *word class*. This means that different value hierarchies can be assigned consistently to different word classes, e.g. nouns vs. pronouns, so that the strategies adopted in order to prioritise information act differently on different word classes, for the same set of category values. Relevant categories are person, gender, case I, and lexicality (the latter is sub-divided in the discussion of the previous sections into subsets, based on word classes). All criteria may give rise to word class conflicts. A different kind of structural conflict is what we call *sub-structure* conflicts. These are conflicts between hierarchies that are related to the same word class, but to different structural components of the class, quite often lexical stem vs. inflection marker. They are found in the categories person, aspect, lexicality, indefiniteness, and among ontological values, and are relevant for the criteria extension, and especially borrowing. This, the relevance of sub-structure divisions to borrowing, is also reflected in the patterns of borrowing discussed in the previous section, where we saw a tendency toward a division between categories that are represented primarily at the level of unbound items, where relevance was a primary motivation for borrowing, and those that are represented mainly at the level of inflection, where the principal motivations for borrowing were transparency and prominence, both related to 'defaultness' and in all likelihood frequency.

Finally, there are conflicts involving the *cross-cutting category*, that is, where the linear order of values or the polarity of a hierarchy will depend on the interaction of two categories. Thus, for example, modals of affirmative ability are more likely to be differentiated than modals of negative ability. The primary category involved is that of modality, and the secondary category is negation. For the criterion of differentiation, the position of the value 'ability' within the modality hierarchy will differ, depending on its interaction

with opposite values from the category of negation (negative vs. affirmative). Cross-cutting category differences are relevant for the categories person, gender, case I, aspect, lexicality, negation, and ontological values, for all criteria except extension.

26.3.2. Conflict categories

Complexity: person. Conflict domain: cross-cutting category, word class. Cf. Section 7.1.

The criterion of complexity reveals conflicting person asymmetries, both in verbs and in personal pronouns. In verbs, the linear order 3–2–1 remains consistent in most cases, and the partial hierarchies do not conflict with it. The only exception to this linear order is the lower complexity of the second person in the imperative (3, 1 > 2 for complexity), which can be explained by the position of the second person as the prototype in the illocutionary domain of the imperative. In all other cases, the second and the first persons show equal complexity in verbs, with the third person being the ambiguous value. In most cases then, we find hierarchies consistent with the linear order 3–2–1, but with conflicting polarity: 1, 2 > 3 or 3 > 1, 2. The former hierarchy may be considered to be the default complexity hierarchy in verbs. The latter hierarchy is exceptional. It reflects the distribution of the perfective intrusion *-in-*, which is more likely to occur in the third person than in the other persons. The domain of occurrence of the intrusion is limited, however, to the the third-person plural, which in turn derives its form from a past participle (adjectival), and is generally more prone to renewal, both internally and through borrowing.

In pronouns, the mutual position of the second and the first persons is ambiguous, and does not appear to be resolvable. The first person tends to be more complex in singular pronouns, while the second person tends to be more complex in plural pronouns. Person complexity conflicts in pronouns are thus determined by the cross-cutting category of number.

Complexity: gender. Conflict domain: word class. Cf. Section 8.1.

While the feminine tends to be more complex in nouns, the masculine tends to be more complex in third-person pronouns and demonstratives. There are no obvious gender asymmetries in complexity in adjectives or verbs. There is no obvious resolution of the conflict. The complexity of the masculine is a result of the tendency toward secondary renewal of the masculine singular third-person pronouns (adoption of demonstrative as anaphoric pronoun), while on the

other hand, in demonstratives, the feminine is more likely to adopt adjectival inflection, simplifying its structure. The greater complexity of the feminine in nouns on the other hand is marginal.

Complexity: lexicality. Conflict domain: structure (word class). Cf. Section 21.1.
The conflict here consists in the greater complexity of lexical verbs over the copula on the one hand (the former are more likely to contain the participial marker *-in-* as a past tense extension), the greater complexity of the less lexical demonstatives over the more lexical adjectives on the other hand (inflectional suffixes of the former tend to be more complex in terms of their phonological make-up), and the greater complexity of the more lexical nouns over the less lexical adjectives (substantival inflection shows greater syntagmatic complexity).

Only a partial resolution of the conflict can be achieved by dividing lexicality hierarchies into sub-sets. Lexicality 1 (auxiliarity) gives the hierarchy non-aux > AUX (i.e. lexical verbs > copula) for complexity. Lexicality 2 (nominal lexicality), which concerns degrees of lexicality in different types of nominals, remains unresolved: the more lexical category is more complex in one instance (nouns > adjectives), but less complex in another instance (adjectives < demonstratives).

Complexity: orientation. Conflict domain: dialect, structure. Cf. Section 18.4.
While in Early Romani the separative orientation was consistently more complex than the stative/directive orientation, various developments have disturbed this asymmetry. Individual dialects show numerous patterns, which may, in addition, differ according to the structure involved.

In pro-words, we find the following dialect-specific complexity asymmetries: (a) separative > stative/directive (pattern retained from Early Romani); (b) separative > stative > directive; (c) directive > separative > stative; and (d) equal degree of complexity of all orientation values. This allows for the generalisation separative ≥ stative, with a volatile position of the directive. In case markers, two conflicting complexity asymmetries are encountered in different dialects: (a) separative > stative/directive (pattern retained from Early Romani) and the opposite (b) stative/directive > separative. There are also dialects where c) both asymmetries are found, though in different localisations: e.g. Sindel *ka* ‘at, to’ vs. *ka-tar* ‘from’ (= a), but *ande* ‘in, into’ vs. synthetic *-tar*/*-dar* ‘out of’ (= b). Similar inconsistencies and conflicts are found in local adverbs (see Chapter 18 for details).

On the whole, there does not seem to be a simple way to generalise over the dialect-specific and structure-specific patterns. The conflicting asymmetries do not appear to be resolvable.

Differentiation: gender. Conflict domain: cross-cutting category, word class. Cf. Section 8.2.
In nouns, the feminine shows more differentiation in number, but the masculine shows more differentiation in inflectional class. Gender asymmetries with regard to inflectional differentiation between different lexical types of adjectivals (viz. descriptive adjectives vs. demonstratives) may assume both directions.

Differentiation: internal case (*case I*). Conflict domain: cross-cutting category, word class. Cf. Section 16.3.
There are conflicting asymmetries between the internal case markers, depending on the cross-cutting category and the word class (Table 26.2). The nominative is more differentiated than the oblique: with regard to number in pronouns and adjectivals; with regard to gender in adjectivals; and possibly with regard to inflectional class in all word classes. The oblique, on the other hand, is more differentiated than the nominative: with regard to number in nouns; and with regard to gender in pronouns.

Let us begin with number differentiation in nouns. The hierarchy by which the oblique is more differentiated relates to the case of generic quantity nouns. Here, there is (optional) number neutralisation in the nominative – thus, SG/PL *řom* '(a) Rom' or 'Roms' – but not in the oblique – SG *řom-es*, PL *řom-en*. Number distinction in the nominative relies on secondary, syntagmatic information, including agreement inflection, which is plural. In pronouns and adjectives, the hierarchy is opposite. In pronouns, there is greater differentiation due to irregularity in the nominative markers *o-v* 'he', *o-j* 'she', *o-n* 'they', compared with the oblique pronominal forms, which carry

Table 26.2. Differentiation hierarchies in the category of internal case

	Number	Gender	Class
Nouns	OBL > NOM	(NOM <> OBL)	(NOM > OBL)
Pronouns	NOM > OBL	OBL > NOM	
Adjectivals	NOM > OBL	NOM > OBL	(NOM > OBL)

the same oblique markers as other nominals: *l-es* 'him', *l-a* 'her', *l-en* 'them'. This is due to the more frequent renewal of the nominative forms, which in turn is triggered by the need to disambiguate salient topics. The original nominative pronouns **lo*, **li* and **le* were thus replaced by the original demonstrative forms *ov, oj, ol/on*. In adjectives, differentiation is greater in the nominative due to greater inflectional potential. The nominative fully differentiates number in most adjectival classes (e.g. M.SG *-o*, F.SG *-i*, PL *-e*), while the oblique shows number neutralisation, or at least a homonymy between the masculine singular and the plural.

Greater differentiation thus attaches to the more salient value, which requires greater disambiguation in discourse. The hierarchies observed in pronouns and adjectives are therefore consistent with the conceptual motivation to prioritise salient or prominent values through greater differentiation. The inconsistency arises due to the behaviour of just a small class of nouns, where the absence of differentiation serves a purpose in obliterating plurality, and so in elevating salient, topical (nominative) generic quantity entities to the immediacy degree that is generally reserved for singular entities. This in itself is, of course, also a prioritisation strategy, one that results in conflict with the strategy of differentiation as applied in other word classes.

The second conflict that we encounter in this domain is the greater differentiation of the oblique in gender marking in pronouns. This results from gender neutralisation in nominative pronouns, but the retention of gender, i.e. of more conservative forms, in the oblique. This is a recent contact phenomenon that is due strictly to Hungarian and to Finnish influence – both languages with no gender differentiaion – on the relevant dialects. Contact influence targets the more prominent, more salient and more frequent nominative forms. This is in line with the overall tendency toward more frequent renewal of the nominative forms in pronouns, and can be explained through the need to reinforce their anaphoric reference function, for purposes of participant disambiguation. The conflict is thus resolvable by disregarding two exceptions for which local motivations can be explained, allowing them to feed into the default differentiation hierarchy nominative > oblique.

Differentiation: aspect. Conflict domain: cross-cutting category. Cf. Section 13.3.

In aspect, perfective forms are more differentiated in evidentiality and inflectional class membership, but less differentiated in number and in person, than non-perfective forms. Evidentiality does not exist in the non-perfective, and there are more perfective inflection classes than non-perfective classes. On

the other hand, concord markers tend to converge in the perfective more often than in the non-perfective.

In terms of the perspective on an event, the perfective is more differentiated than the non-perfective, due to its focus on the result. Here, it is more important to stress the effect of the underlying action or event on the actor/undergoer. This explains the presence of evidentiality as a statement about the source of knowledge (cf. Matras 1995), and the semantic differentiation of inflection classes relating to the agentivity status of the subject. With respect to subject concord marking, or the identification of the actor/undergoer that is involved in the action, the perfective is more likely to undergo levelling of number and person in particular in the second person. In the non-perfective, which enjoys greater accessiblity of the action/event and so also of the principal participant in it, such levelling is more likely to be constrained. Although both hierarchies can be explained as outcomes of local motivations, there appears to be no convenient way of reconciling them and resolving the conflict.

Differentiation: negation. Conflict domain: cross-cutting category. Cf. Section 10.2.

On the one hand, negated verbs tend to be more differentiated in TAM categories than affirmative verbs. On the other hand, modals of affirmative ability are more likely to be more differentiated than modals of negative ability. In Kalburdžu, for example, we find the forms ***ma*** *ker* 'do not do!', ***in*** *ker-e* 'you are not doing', and *te* ***na*** *ker-e* 'that you (may) not do' – displaying different negators in different TAM categories. If we take ability and inability expressions, we find that it is the affirmative expression that is more likely to be differentiated, inflecting for person, number, and TAM, while in the negative, the uninflected and impersonal particle *našti* 'cannot' is more likely to be retained. The latter, the tendency to retain *našti*, is relative, of course, to the non-retention or weaker retention of its affirmative counterparts *ašti* or *šaj*. The latter, the affirmative, is simply more frequently replaced by borrowings, which in turn is well in line with the overall tendency to borrow those inflected expressions that are more salient, frequent, and higher on the truth and simplicity and transparency scales. In lexical verbs, we do not encounter a similar hierarchy, since there are no parallel cases where the affirmative expression is borrowed, but the negated expression is retained.

This highlights a methodological difficulty. The grammaticalised expressions of ability and inability are indeed tokens of two polar values in the category paradigm of negation. However, with lexical verbs, the motivation to differentiate the negator rests not within the paradigm, but it is triggered at

the level of the syntactic environment. This shows that a strictly paradigmatic approach to asymmetry has its limitations. The default hierarchy (affirmative > negative, for differentiation), which is in line with the greater differentiation of values that are high on the prominence, the truth and simplicity, and the transparency scales, is preserved for modals. The other, conflicting hierarchy (negative > affirmative, for differentiation), is an exception that is motivated by syntagmatic factors.

Differentiation: lexicality. Conflict domain: word class, cross-cutting category. Cf. Section 21.2.
Pronouns are more differentiated than nouns in that they are more likely to retain synthetic case inflection; in that they exhibit a greater stem differentiation; and in that they are more likely to have irregular and differentiated genitive (possessive) forms. However, pronouns are less likely to retain gender distinctions than nouns. The loss of gender in pronouns is a case of contact influence on nominative pronouns (see above). They are much more frequent than nouns, and are also more prone to renewal than nouns, especially when we consider that gender reduction in pronouns only occurs in the more salient and frequent nominative. In a sense, inflectional reduction in the nominative pronouns appears counterfunctional, since one would expect a greater need for disambiguation in nominative pronouns. On the other hand, it is clear how contact influence in the area of inflection makes its way into more salient entities first. This illustrates just how spectacular contact-induced change is: the functionality of change in reaction to contact – lifting some of the demarcations between sets within the bilingual's linguistic repertoire – is not always in line with the functionality of information organisation in monolingual discourse. We have, in other words, a clash of motivations, leading to conflicting strategies of prioritising information in respect of category values. This in turn results in a conflict between asymmetry hierarchies.

The copula tends to be more differentiated than lexical verbs, showing a greater differentiation in terms of TAM distinctions, inflectional irregularity, and through its greater propensity to co-occur with subject clitics. Similarly, pronouns are – prevailingly (see above) – more differentiated than lexical nouns. On the other hand, substantivals are more differentiated than adjectivals; and descriptive adjectives and cardinal numerals tend to be more differentiated than less lexical modifiers. This conflict can be resolved by splitting lexicality into two sub-categories (cf. Chapter 21). Auxiliarity (= lexicality 1), in which copula and pronouns are treated as auxiliaries, and lexical verbs and nouns as non-auxiliaries, has the differentiation hierarchy auxiliary >

non-auxiliary. Nominal lexicality (= lexicality 2), consisting of various modifiers and nouns, shows the differentiation hierarchy noun > modifier (adjective > possessive > article).

Extension: gender. Conflict domain: dialect, word class. Cf. Section 8.3.
Gender extension is only found in the inflection of adjectivals (including specific developments in demonstratives and the definite article), and in the inflection of third-person pronouns. The developments in adjectivals suggest that masculine forms will extend to the feminine, rather than vice versa. This is, however, not confirmed by the developments in personal pronouns, where both directions of gender extension are attested. The developments in pronouns, however, are the ones triggered by contact with genderless languages. It therefore appears that, in the neutralisation of gender through convergence with a genderless system, the form that prevails is chosen at random, and not as a specific, meaningful value of the category paradigm of gender. The exception to the hierarchy therefore does not question the default extension hierarchy, masculine > feminine.

Extension: aspect. Conflict domain: dialect, substructure (stem vs. adaptation marker). Cf. Section 13.3.
Aspect extension is attested in both directions. In Early Romani and in most dialects, there are a few verbs that show an irregularity in the formation of their perfective stems. While most verbs derive their perfective stem by suffixation of a perfective marker to the non-perfective stem (e.g. NPFV *ker-* > PFV *ker-d-* 'do'), some irregular verbs in addition undergo an irregular stem alternation (e.g. *mer-* > *mu-l-* 'die') or even suppletion (cf. *dža-* > *ge-l-* 'go'). Some dialects have abandoned this irregularity by extending, optionally or obligatorily, the non-perfective stem to the perfective forms in some or all of the irregular verbs. Here, therefore, the non-perfective is more likely to extend.

Cross-dialectal evidence suggests that, in Early Romani, adaptation of borrowed active verbs involved the Greek suffixes *-Vz-* or *-Vn-* in the non-perfective, and the Greek aorist suffix *-Vs-* in the perfective (cf. Matras 2002: 128–133). In many dialects the non-perfective adaptation markers have extended to the perfective (e.g. East Slovak Romani *us-in-d-e* 'they swam', Kosovo Bugurdži *piš-iz-d-e* 'they wrote'), while in most Vlax dialects and partly also in Welsh Romani, the perfective adaptation marker has extended to the non-perfective (e.g. Kalajdži *vorb-is-ar-en* 'they speak').

In adaptation markers, the distinction between perfectivity and non-perfectivity is not unambiguous. Modelled on the intransitive derivation of verbs

from transitive bases by adding a valency-decreasing marker to the perfective transitive stem (*ker-d-jov-* 'to be done'), intransitive loans are often adapted on the basis of a Greek-derived aorist marker: *jir-is-áv-av* 'I return', from Greek aorist *jir-is-* 'to have returned'. The original Greek aorist form thus becomes a non-perfective, and the perfective is formed by adding a perfective marker: *jir-is-aj-l-jom* 'I returned'. Now, the perfective adaptation marker that extends into the non-perfective in Vlax and Welsh Romani is this same Greek-derived aorist *-is-*, which, as we said, is not unambiguously perfective due to its presence, already in Early Romani, in non-perfective intransitives. The conflict is thus resolvable, with the exception feeding into the default extension hierarchy non-perfective > perfective.

Extension: cardinality. Conflict domain: dialect. Cf. Section 11.3.
Both directions of the extension of additive connectors – from lower to higher ten+unit numerals and vice versa, and from lower to higher ten numerals and vice versa – are attested, although we are not able to evaluate their relative frequency.

Borrowing: lexicality. Conflict domain: word class, substructure (stem/word vs. inflection). Cf. Section 21.4.
We have already seen above that different rules are at work in the borrowing of inflection, and of lexical items. Lexical roots are more likely to be borrowed than roots with grammatical function (modals, copula, pronouns), showing lexicality as high on the borrowing scale. Inflection, however, is more likely to be borrowed with borrowed modals than with borrowed lexical verbs. This can be explained as a complete fusion of the operational procedures that modify propositions, a process that, like any fusion in a selective domain of utterance organisation, is more likely to occur than either the wholesale or selective replication of predication-anchoring morphology – verb inflection – with lexical items. The transfer of inflection in modals allows modals to constitute a sub-group based on their grammatical function (though compartmentalisation within the lexicon is also known for Turkish verbs in some Romani dialects of the Balkans). The hierarchy of inflection borrowing is not, however, entirely inconsistent with that of lexical roots if one takes into account that the transfer of the lexical root in modals is a pre-requisite for the transfer of inflection markers; thus, the borrowing of lexical roots precedes that of inflection. The exception therefore feeds into the default hierarchy more lexical > less lexical, for borrowing, and specifically to that of auxiliarity, non-auxiliary > auxiliary.

26.3.3. Conflict pairs

Complexity: tense: FUT – IMPF. Conflict domain: dialect. Cf. Section 13.1.
In those dialects where the future is encoded by the long verb forms, the imperfect is more complex than, or at least as complex as, the future. However, in dialects that possess an analytic future and, at the same time, the original synthetic remoteness marker, it is the future that is more complex than the imperfect. In some dialects both categories are encoded analytically, and there is no obvious asymmetry in complexity.

Complexity: modality: ability – necessity. Conflict domain: dialect. Cf. Section 14.1.
Table 26.3 shows the presence/absence of the complementiser *te*, in different dialect types, that may link modals (verbs or uninflected particles) with their complements. Only dialects with complexity conflict between the ability modal and the necessity modal are shown. The type numbers refer to the complete Table 14.2 in Chapter 14.

In Types 3 and 7 the ability modal requires the complementiser, while it is only optional with the necessity modal. On the other hand, in Types 4 and 9 the necessity modal requires the complementiser, while it is only optional or absent with the ability modal.

Complexity: ontological: determiner – others. Conflict domain: dialect, cross-cutting category (deictic: interrogative vs. indefinite). Cf. Section 20.1.

The determiner is the most complex value in interrogatives and in indefinites based on interrogatives, while it is the least complex value in indefinites with generic bases. Interrogative determiners (e.g. *s-av-o* 'which') consist of an interrogative root, an ontological marker, and an adjectival inflection through which they agree with their head nouns. The other interrogatives, on the other hand, are either indeclinable (consisting of an interrogative root and an ontological marker, e.g. *s-ar* 'how') or they tend to have cumulative encod-

Table 26.3. Complementiser te in selected modal complements with identical subject

Dialect	'can'	'must'
Type 3 & 7	te	–/te
Type 4	–/te	te
Type 9	–	te

ing of ontological and (substantival) inflectional marking (e.g. *k-on*, OBL *k-as* 'who'). The complexity of the determiner derives from the need to accommodate agreement marking. The aforementioned complexity asymmetry is retained in indefinites based on interrogatives (e.g. *vare-s-av-o* 'some' > *vare-s-ar* 'somehow').

In contrast, the determiner is the least complex value in indefinites with generic bases. Indefinites of most ontological categories employ generic nouns as their ontological markers (e.g. Sepečides *hidžek dženo* 'nobody', lit. 'no person'*, hidžekhe thaneste* 'nowhere', lit. 'at no place'). The indefinite determiner, on the other hand, accompanies the noun it determines, and so its ontological category does not need to be further specified by a generic noun (e.g. Sepečides *hidžek kher* 'no house'). The determiner thus consists of an indefiniteness marker alone, without containing an overt marker of the ontological category.

The conflict derives from a different choice of strategies in encoding ontological distinctions in indefinites in different dialects. One of the strategies is to draw on the set of interrogatives, which provide an off the shelf grid of ontological marking. The other strategy is to construct ontological distinctions independently of interrogatives, by means of generic nouns. The conflict is thus explainable, but not resolvable. If we were to separate the strategies into two categories – that is, define an 'interrogative–ontological' and a 'generic–ontological' category – we would be creating in effect sub-samples, rather than attempting generalisations about the categories that are naturally reflected in the sample as a whole.

Differentiation: person: 1, 2 – 3. Conflict domain: cross-cutting category. Cf. Section 7.3.

The first person is always more differentiated than the second person. The conflict concerns the hierarchical position of the third person: it is the most differentiated value in the cross-cutting categories of gender, case, and class, but not in the cross-cutting category of number. Also, the third person assumes conflicting hierarchical positions with regard to differentiation of TAM categories.

The person hierarchy of number differentiation in pronouns is 1 > 2 > 3: first-person pronouns exhibit strong number suppletion of their roots (cf. 1SG *m-* vs. 1PL *a-*), while the roots of second- and third-person pronouns are not suppletive across number. Unlike the third-person pronouns, the second-person pronouns show some irregularity in their stem formation. Also, first- and second-person pronouns employ an irregular plural marker *-m-*, while in third-

person pronouns number marking is irregular in the nominative, but completely regular in the oblique forms. We are dealing here with depth of differentiation, rather than the presence or absence of the number distinctions as such, which remains unaffected. There is therefore no functional reduction in the lower positions, but simply more 'tolerance' of irregularity in the higher, first- and second-person positions, which we may attribute to egocentricity.

In verbs, the person hierarchy of number differentiation is 1 > 3 > 2, reflecting the possibility of number homonymy in the second and third persons, but not in the first person. However, number homonymy in the third person is marginal: it is usually restricted to very specific TAM contexts; or it only concerns inflections rather than word-forms; or it is licenced by second person-homonymy. Here, the need to disambiguate third-person referents is presumably a weaker motivation to differentiate number than the egocentricity of the first person. Since second-person referents are in a lesser need of disambiguation and, at the same time, the second person shows only medium degree of egocentricity, number homonymy in the second person is tolerated in some dialects.

As for TAM differentiation, the third person is the most differentiated value whenever one looks at distinctions of morphosyntactic nature. On the other hand, the third person is the most likely value to develop TAM homonymies through phonological erosion. Here, the greater need to disambiguate third-person referents is reflected in asymmetries of morphological origin, while the greater susceptability of the third person to erosion effects is presumably motivated by its higher frequency.

Differentiation: ontological: determiner, person, thing – place, quantity. Conflict domain: cross-cutting category.

Place pro-words are highly differentiated by the cross-cutting category of orientation, and quantity pro-words show differentiation into cardinals, ordinals, and multiplicatives (e.g. Slovak Romani *ajci* 'that much', *ajci-to* 'in that place in an order', and *ajci-var* 'that many times'). The criteria of declinability and inflectional differentiation thus render the following ontological asymmetry in pro-words: person > thing > ordinal quantity > determiner > cardinal quantity > place, time, manner, cause/goal, multiplicative quantity. This highlights the specific difficulty associated with the criterion of differentiation: it relies on cross-cutting categories. Depending on the main category, individual values may or may not logically allow or attract certain cross-cutting categories. In this case, place words are sensitive to a different cross-cutting category.

Extension: person: 1, 3 – 2. Conflict domain: word class, dialect. Cf. Section 7.4.
In verbs, second-person forms may extend to the third person – cf. perfective 2PL *-an* > 3PL *-e* giving 3PL *-en* – and vice versa, third-person forms may extend to the second person – 3PL *-e* > 2PL *-an* giving 2PL *-e*. If the first person is affected, then the extension proceeds along the scale 3 > 2 > 1. In other words, verbs show an extension 2 > 1 only if there is also an extension 3 > 2; and so 2 > 1 is a completion of a step towards uniformity of the vowel pattern throughout the person paradigm. The vowel pattern then becomes a number marker in the perfective: Ukrainian Romani for instance has *-e* in all plural forms, and Abruzzian Romani plural forms all contain the vowel *-e-*: 1PL *kerdem*, 2PL *kerden*, 3PL *kerde*. In this respect, there is a shift between the categories person and number, as number is taken out of the cumulative expression, with priority given to the uniformity of number.

In pronouns, on the other hand, first-person forms may extend to second-person forms, as in the possessives *minřo* 'my' > *tiro* 'your' giving *tinřo*. The analogy in pronouns is based entirely on egocentricity, while verbs show greater flexibility of motivations, to the extent that the two possible directions of analogy of verbs constitute a non-resolvable, dialect-based conflict of hierarchies.

Borrowing: person: 1, 2 – 3. Conflict domain: substructure (person marker vs. number marker). Cf. Section 7.7.
The Greek-derived 3SG marker *-i* is the most frequently attested borrowed person marker. Quite possibly it entered the language initially with impersonal modals; a similar development is underway at present in some dialects (cf. Slavic *može* and Greek *prepi* in some dialects of the Balkans). The borrowing of person markers (3SG > 2PL > 1PL, 1SG) from Slovene into Slovene Romani is inconclusive. For a start, there is contamination of forms in the first person. Moreover, since Slovene Romani generalises Greek-derived *-i* in the 3SG, there is also no conflict between the borrowing of the third-person singular at the first stage, and only subsequent borrowing of other person forms.

The actual conflict is with borrowing of Turkic-derived number markers into different persons of the perfective inflection of non-Turkic verbs: they are more likely to be borrowed into the second person than into the first person, and they are never borrowed into the third person. But the motivation here is different: borrowing of number markers is inspired by the similarity between the singular markers, which in both languages show *-m* in the first person and *-n* in the second person. Symmetry is obtained by adding Turkish plural mark-

ers to the Romani person inflections. This analogy is absent in the third person. Quite possibly, there is even an additional motivation: If the hierarchy of borrowing from Turkic is 2 > 1, then the borrowing of the Turkic number marker is functional in the second-person perfective in order to distinguish plural and singular, where the relevant dialects of Romani used to have *-an* for both. This of course presupposes that borrowing occurred prior to the form extension from 3PL *-e* to 2PL *-en* (see Chapter 7), thereby distinguishing 2SG *-an* from 2PL *-en* – a developed which later took place in most dialects of the Balkans. There is clear evidence, however, to support the assumption that the borrowing of the Turkic markers occurred before this analogy (see Section 7.7, cf. Ajia Varvara 2PL variants *-en* and *-an-us*).

From this we may conclude that there is a default hierarchy 3 > 2, 1 for the borrowing of person markers, with just a very specific local motivation to borrow plural markers according to a different person hierarchy.

Borrowing: indefiniteness: free-choice – universal. Conflict domain: substructure (affix vs. word). Cf. Section 19.5.
The criterion of borrowing renders two partly conflicting asymmetries, depending on whether one considers borrowing of indefiniteness markers (1) or borrowing of whole indefinite word-forms (2).

(1) **Free-choice** > negative > specific > **universal**
(2) Negative, **universal** > specific > **free-choice**

The position of free-choice indefiniteness in the two hierarchies is reverse: free-choice markers, such as *bilo-*, *-godi*, *ori-*, whose meaning corresponds to 'any- whatsoever', are the most likely to be borrowed; but free-choice indefinites, with meanings such as 'anywhere whatsoever', are the least likely to be borrowed. A similar pattern holds for universal indefiniteness: universal markers with the meaning 'every-' are not likely to be borrowed, but universal indefinites with meanings like 'everywhere' are very likely to be borrowed.

In respect of the universal value, there are two conflicting, but complementary tendencies. The first is to borrow the universal function – among the indefinites – as it is high on the relevance scale. We thus find numerous borrowings for expressions like 'always': *mindig* (Hungarian), *zavše* (Polish), *imer* (German), and so on. The second tendency, is to use whole word-forms for the universal, rather than employ a combination of marker with ontological specifier. With the free-choice indefinite we also find two complementary tendencies. The first is to borrow the free-choice function, for it too is high on the rele-

vance scale. The second is to use transparent free-choice expressions.

Thus, with the universal, if a full word form is not available as a potential borrowing, then the composition of a universal is also avoided. This places universal *markers* low on the hierarchy, since, although the function itself is highly borrowable, there is a preference for whole word forms rather than analysable compositions. If the tendency to avoid universal compositions is shared by the contact languages, then there would also be few analysable universal markers on offer as potential borrowings. If we now turn to the free-choice markers, we see that, for a start, not all dialects (or languages) will distinguish them from universal expressions. If they are distinguished, then we find an overwhelming tendency to construct them consistently, using analysable compositions, rather than word forms. This increases the demand for a free-choice marker, which may derive either from within the inherited inventory, or from the inventory of the contact language. If a similar tendency is common in the contact languages, then this would of course increase the availability of analysable free-choice markers that can be adopted as potential borrowings.

If we now turn to the contact languages, we find that these predictions are precisely met. The West and East Slavic contact languages – Czech, Slovak, Polish, Russian – all have transparent free-choice expressions, derived from the respective interrogatives as markers of the ontological categories, with addition of a free-choice marker. By contrast, the universal set is irregular and less transparent. Borrowings into Romani are the transparent free-choice markers, and the less transparent word forms for universal expressions. In German, there is no opposition between universal and free-choice function, and hence no separate set of free-choice expressions or free-choice markers. As a result, only universal word-forms are borrowed into Romani. In Hungarian, the free-choice set is marked by *akár-* and so consistently transparent. The universal set includes *mind-*/*minden-*; note already the variation in the marker. In addition, the ontological element of the universal expression is not consistently derived from the interrogative. The universal set is therefore much less transparent. Turkish, another agglutinative contact language, has transparent and analysable markers in both sets, but the free-choice expressions are more complex, thus confirming the tendency to construct free-choice expressions through composition: cf. *her yerde* 'everywhere' (universal), and *herhangi bir yerde* 'anywhere whatsoever' (free-choice). In Greek, we find the marker *dhipota* in the series for free-choice expressions, and so a transparent structure, while universals are lexicalised: e.g. *panda* 'always', *kathenas* 'everybody', *pandou* 'everywhere'. Finally, in Rumanian, *ori-* functions as a free-

choice marker, while universal expressions are, similarly, lexicalised: e.g. *tot* 'every', *mereu* 'always'.

In conclusion, we find a conflict that is not, perhaps, resolvable, but is explainable at the local level: Indefinites are generally borrowable, which is a function of their position on the relevance scale (demanding intense mental cooperation from the hearer in order to locate information based on presuppositions). There is, however, a hierarchy of borrowability. The values free-choice and universal assume opposite positions on that hierarchy; moreover, their positions change according to whether the hierarchy is applied to refer to the borrowing of indefiniteness markers, or entire (lexicalised) word forms. This clash is an accommodation to universal tendencies of constructing the two values, whereby free-choice expressions are more likely to be composed and so transparent, while universal expressions are more likely to be lexicalised. This tendency is also found throughout the contact languages. With the free-choice value, therefore, borrowing of a marker is more likely, while with the universal value, borrowing of the word form is more likely.

26.4. Motivations for asymmetry: Concluding remarks

Our point of departure when approaching the data was that asymmetry in language results from the application of a number of strategies at the local level, in order to help prioritise information, based on conceptual categorisations inspired by cognitive categorisations of real-life. From the point of view of the efficiency of communication, several factors trigger the need to prioritise information: saliency of the topic of talk; the degree of transparency of entitites; the accessibility of information in discourse; the relevance of information to discourse presuppositions; and egocentricity or the position of the speaker relative to the object of talk. Information is categorised, drawing on analogies to and representations of real-life entities and the relations among them: person, gender, degree, valency or discreteness are just a few examples. Within these categories, values are differentiated, again by analogy to the categorisation of entities in the real-world (action upon another object vs. action affecting the actor, as an example), or indeed constituting a speech-based categorisation of the real world (e.g. the system of cardinality). Prioritisation is applied arranging these values in a hierarchical order, and by drawing on a pool of available strategies that shape the structure of language, in a way that favour one pole in the hierarchy over another. The Markedness Hypothesis predicts a direct relationship between the pole that is favoured, the nature of

the strategy, and the relationship between values and what they represent in the real world.

We have found some confirmation of this hypothesis, in the overwhelming tendency of language structures to be shaped in an asymmetric way, revealing prioritisation of certain category values over others in communication. Moreover, we have seen that language change is guided overwhelmingly by such principles of asymmetry and hierarchical relations within categories, and that related speech varieties therefore tend – with few and rather marginal exceptions – to show similar hierarchies. In the present chapter, we identified tendencies toward grouping of categories, into iconic representations of several dimensions, including quantity, immediacy, prominence, transparency and truth and simplicity. Though the borderlines between these groupings are not always robust and clear-cut, there are tendencies for certain dimensions to favour certain prioritisation strategies. Moreover, these strategies tend to cluster in different ways for different groups. This confirms our assumption that markedness is not a pre-determined cooccurrence of criteria, forged by a pre-determined cooperation of strategies. Rather, markedness theory must be understood as the more abstract principle governing the tendency to shape language asymmetrically, and in relation to cognitive categorisations. How and through which strategies asymmetry is created, will depend on information structure and its utilisation in communication, on its iconic presentation, and on the values used to represent analogies between pieces of information and real-world conceptual entities. That local motivations of this kind are not always predictable, but will compete among themselves, was illustrated in detail in the final part of this chapter.

Chapter 27
Concluding remarks

Our point of departure in this book has been that language shows a competition between two tendencies: one toward consistency and so toward a form of symmetry, the other toward prioritisation of certain category values over others, and so toward asymmetry. We have attempted to carry out an exhaustive survey in related varieties of an inflected language, in order to establish what are the prevalent patterns that are represented by the variation of structures among the dialects of Romani, and by inference, the nature of the processes of change that led to this variation. The most striking outcome of our survey is the prevalence of asymmetry. For most of the categories that we examined, and by most of the criteria which we applied, paradigms are usually structured in an asymmetrical way. The instances of symmetrical structuring are few and on the whole marginal.

Although asymmetry is prevalent, it is not always represented by a neat, smooth pattern. Asymmetry hierarchies are complete and consistent only for some categories. For some, only partial hierarchies can be extracted as consistent. Others show conflicting hierarchies. The overwhelming tendency, however, is toward an identifiable linear ordering of values within a category for a given criterion; there are, in other words, relatively few cases in which a single criterion renders conflicting hierarchies. Even in those categories for which conflicting linear ordering of values is found, there is a tendency toward a hierarchisation of the conflicting orders to form a dominant order pattern, and one that is aberrant or subordinate. The first – dominant – order will attract the majority of asymmetry criteria, while the other – subordinate – order will follow just a minority of criteria. The Romani sample thus provides us with the impression that language structures serve to prioritise information following a rather strict prominence order of values within individual categories.

Not all criteria for asymmetry, however, are as prominently represented across the categories. The criteria that are relevant to the largest number of categories are ‘complexity’ and ‘differentiation’ (both appearing in 22 different categories). These can be regarded as major strategies used in order to prioritise information. They are followed by criteria of intermediate relevance, ‘exposition’ (13 categories) and ‘extension’ (12 categories). Of marginal relevance are the criteria of ‘extra-categorial distribution’ (7 categories)

and 'erosion' (6 categories). Alongside these criteria, borrowing is of major importance in Romani as a means of prioritising category values, and is relevant to no fewer than 19 different categories. Internal diversity, by contrast, is relevant only to 12 categories.

Our principal aim in this investigation was to test – for a sample of closely related varieties and as exhaustively as possible for a range of different categories represented in different linguistic structures – what we referred to in the introductory part as the 'Markedness Hypothesis': the assumption that there is a tendency in language to treat paradigmatic values as 'marked' or 'unmarked'. This assumption entails a prediction about the way in which criteria will cluster for a given value among a pair, more specifically about the polarity of value ordering. Thus, the typical 'unmarked' value is (–)complex, (+)differentiated, (+)eroded, and so on. In practice, then, testing the Markedness Hypothesis means checking whether the direction or polarity of value ordering for the main criteria is in line with this prototypical characterisation of 'marked' and 'unmarked' values.

Our findings point to the partial truth or reality of the Markedness Hypothesis: There are categories, albeit few, that appear to conform to the prediction, and which we have labelled 'well-behaved categories'. A well-behaved category by our definition is one that attracts a cluster of at least four criteria, indicating that considerable effort is made, drawing on a variety of different strategies, to prioritise values following a consistent pattern of value ordering. Language appears – to judge by our sample – to put this effort into a type of category in which one side of the value hierarchy – the 'marked' side – tends to be associated with conceptual complexity, while the other – the 'unmarked' – side is associated with the mental accessibility of conceptual values. The few categories that meet this characterisation are Number, Degree, and the ontological pair Thing–Cause. A further set of categories – among them Gender and Tense – may be said to have default values that are 'unmarked' by a cluster of criteria that includes 'extension', 'exposition', and 'extra-categorial distribution'. One of our major findings, then, is that well-behaved categories, and even categories that follow the predictions of the Markedness Hypothesis in a somewhat more remote sense, account for just a fraction of the asymmetry hierarchies attested in the sample. To reiterate, then, there is some reality to the notion of markedness, but only in a small number of categories, and in relation to a relatively modest set of criteria.

The clustering behaviour of both categories and criteria is the key to any overall picture of asymmetry. Categories cluster in the sample by displaying matching sensitivities to groups of asymmetry criteria. We were able to

identify several different clusters, ranging from the 'well-behaved' categories, whose 'unmarked' values are prioritised by appearing formally as simplex (non-complex), unique (exposed), multifaceted (differentiated) and buoyant (prone to extension), and which represent conceptual values that are easily accessible and cognitively simplex, to those that single out conceptually independent and transparent values as 'marked' by prioritising them through greater formal complexity, greater differentiation, uniqueness as well as buoyancy.

We have used the term 'iconicity' in this connection to refer to the relation between the bundle of formal strategies that are employed in order to prioritise values, and the conceptual properties that are associated with those values across the different categories. Values representing higher quantities are more complex (being more difficult to process), less differentiated (being more difficult to pinpoint in reference to other categories), less unique (expressing generalisations across single entities) and less buoyant (less prominent as metaphors). Less immediate values (e.g. remoteness, future, separative) are more complex (being less easy to grasp conceptually), more exposed (requiring more obvious identification), and less differentiated (playing less of a role in the processing of salient events and so requiring little explicit interplay with other categories). More independent and transparent values are both more complex and more differentiated, representing a greater number of contextual environments and potential for interplay with other categoric values.

Apart from their clustering behaviour, the representation of categories in the sample provides us with insights as to the overall prominence of those categories in the language: Those that attract a greater number of asymmetry criteria can be said to be more frequently targeted for asymmetry, in other words, they constitute the categories for which greater effort is being made to prioritise values. Defined in this way, the more prominent categories are number, degree, person, gender, discreteness, case, tense, transitivity, indefiniteness, and the ontological 'thing' value.

Contrary to the Markedness Hypothesis, there is no consistent link between the major criteria pair 'complexity' and 'differentiation', nor between the pair 'complexity' and 'exposition'. The criterion of 'erosion', itself rather marginal, cannot be linked consistently to either 'differentiation', 'complexity', or 'extension'; we can thus find no evidence that proneness to erosion is a property that accompanies simplex, default or otherwise defined 'unmarked' values. General proneness to renewal, measured by the criterion of internal diversity, can be linked with structurally complex values that share the property of con-

ceptual complexity (such as plurals, superlatives, or high numerals), as well as with structurally simplex values representing salient or default entities as well as citation forms.

Romani is a language in contact, and borrowing figures prominently among the strategies used to discriminate between paradigm values. But no plain correlation can be found between borrowing, and any particular relation among the major asymmetry criteria 'complexity' and 'differentiation', and therefore no correlation between borrowing and any conventional notion of 'markedness'. Borrowing may target values representing higher quantity, lower saliency and weaker accessibility, all of those 'marked' values by the conventional notion (greater complexity, weaker differentiation); or alternatively, it can target values of greater independence, greater prominence, and presumably higher frequency, which satisfy the conventional criteria for 'unmarked'. It appears then that the motivations for borrowing are not directly linked to the language-internal strategies of prioritising information. Rather, borrowing serves to support speakers in managing choices among the structures in their bilingual repertoire. Specifically, borrowing is the speakers' licence to eliminate the choice, by allowing the two systems to merge around a particular structure. Such a licence is most functional around elements that are high on the relevance scale and demand greater attention to the hearer's expectations – often elements that show high complexity; as well as around elements that are stereotypically salient and are employed in order to accommodate lexical borrowings – usually those that are more frequent, more default, and less complex. In this respect, borrowing may be said to target both 'marked' and 'unmarked' values.

This book presents a discussion of Romani in its (considerable) dialectal variation. Whilst there is no reason to assume that Romani is unique and exceptional in the way it treats conceptual notions and maps them onto formal strategies to prioritise information in discourse, it would be naïve, as well as empirically unfounded, to assume that what is true for Romani is typical for most other languages, and it would be premature to view our conclusions as representing universal properties of language. Some of the categories investigated here have not, to our knowledge, been investigated in a cross-linguistic sample. For some others, some of our observations match those already made on the basis of a broader language sample. Our goal was to demonstrate exhaustively the role of asymmetry in the patterns of a language – a task which has not hitherto been attempted, as far as we are aware. The outcomes will need to be tested against a representative sample of languages in order to ascertain their validity as generic properties of the language faculty.

Appendix I: Sample dialects

This section provides an overview of the Romani dialects of our sample, classified into 12 genetic groups (cf. Matras 2002). There is information on the dialects' location, contact languages, source of data, notes on internal subgrouping within a dialect group, and notes on alternative dialect names. All dialects share Greek and South Slavic as early contact languages, and so only later contact languages are indicated. Data gained through the Support Elicitation Questionnaire are marked as "RMS (Support Elicitation)". Unpublished data from the RMS database which have not been gained through Support Elicitation are marked as "RMS (name of fieldworker/inputter)".

If not otherwise mentioned, the primary ethnonym of the speakers is *Řom* or *Rom*. The names of the dialects are those used in Romani linguistics, and they do not necessarily reflect the names employed by the speakers. Some of the dialect names merely indicate their location (e.g. *Kaspičan* is the Romani dialect of Kaspičan, Bulgaria), others reflect the group's name (e.g. *Zargari*). Occasionally, we have to combine these two in order to distinguish distinct dialects of speakers who use an identical group name (e.g. *Montana Kalajdži* is a Balkan dialect, while *Varna Kalajdži* is a South Vlax dialect). Dialect names based on names of countries are conventional rather than descriptive (e.g. *Polish Romani* is only one of three major dialects spoken in Poland, and *Slovene Romani* is also spoken in Italy). Some of the 12 groups given below can be lumped together: the "northern" dialects comprise the British, Northwestern, and Northeastern dialects; the *Central* and *Vlax* dialects both consist of the respective North and South (sub)groups; and the *Balkan* dialects consist of the South Balkan dialects and the Balkan *zis*-dialects. We distinguish Balkan dialects from dialects "of the Balkans": the latter comprise the South Vlax group as well as the Balkan groups.

British

There are records of two dialects of the British group: English Romani spoken by the *Romanichals* in England and other parts of Great Britain, and Welsh Romani spoken by the *Kāle* in Wales. Both dialects are now extinct. English Romani was recorded at a progressed stage of linguistic obsolescence (accompanied by a shift into a Romani

Table I.1. British dialects of Romani

Dialect	Location	Source
English R	Great Britain (England)	Smart and Crofton 1875
Welsh R	Great Britain (Wales)	Sampson 1926

ethnolect of English). Welsh Romani, on the hand, was richly documented as spoken by fully competent speakers in the beginning of the twentieth century. The current L2 of English Romani was English and the current L2s of Welsh Romani were English and Welsh. Previous L2s included French, German, and West Slavic.

Northwestern

The Northwestern group is divided into two internally homogenous subgroups: Finnish Romani[1] and Sinti. Finnish Romani is a special case. The language of the *Kaale*, who live in Finland and Sweden, is used alongside Finnish by the older generations only, and so arguably it is no longer being acquired as a first language: modern Finnish Romani "is not learned primarily in childhood, but gradually as Roma children are introduced into the life and activities of adulthood, where the language is used as a secret language" (Borin 2000: 75). However, it remains the community language in a symbolic sense (thus Romani language instruction in Finland is officially referred to as 'mother tongue education'). Accommodated onto our normal grid of contact languages, Finnish is the current L2 (the second language actively spoken by the entire community alongside Romani), Swedish the recent or current L2 (Swedish was the major L2 for most of the Romani population of Finland several generations ago, and is now of course a second L2 for the Finnish Roma who have emigrated to Sweden), and German an older L2. The Sinti dialects are spoken by the *Sinti* (*Cinti*) of Germany, Czech Republic, Austria, Hungary, Italy, France, and other European countries, and by the *Manuš* of France. We use the term *Core Sinti* to exclude the slightly aberrant Piedmontese Sinti. The current L2s are those of the respective countries (although German is also partly retained as a current L2 e.g. in Hungary), with German being the formative past L2 of all Sinti dialects.

Table I.2. Northwestern dialects of Romani

Dialect	Location	Source
Finnish R (older)	Finland	Bourgeois 1911; Valtonen 1972
Finnish R (various modern)	Finland	RMS (Support Elicitation)
Sinti: German (older)	Germany	Finck 1903
Sinti: German (modern)	Germany	Holzinger 1993
Sinti: Lalere	Germany, Czech Republic	Holzinger 1993
Sinti: Austrian	Austria	RMS (Heinschink and Schrammel)
Sinti: Hungarian	Hungary	Mészáros 1980
Sinti: Manuš	France	Valet 1991
Sinti: Piedmontese	N Italy, S France	Franzese 1985
Sinti: Lombardian	N Italy	Soravia 1977

Table I.3. Northeastern dialects of Romani

Dialect	Location	Source
Polish R	Poland	RMS (Support Elicitation), Matras 1999a
Podolie R	NW Ukraine	Barannikov 1934
Lithuanian R	Lithuania	RMS (Support Elicitation)
Latvian (Curland) R	Latvia	Mānušs, Neilands, and Rudevičs 1997
Russian R	Russia	Wentzel 1980
Estonian R	Estonia	RMS (Support Elicitation)

Northeastern

The relatively homogeneous Northeastern dialects are spoken in Poland, Lithuania, Latvia, Estonia, Belarus, northwestern Ukraine (the Podolie region), and most parts of Russia and some other CIS republics. Significant communities of recent outmigrants live in Britain and Germany (Polish Romani), and possibly still in China (Russian Romani). The Northeastern dialects share German as an older contact language. Polish is the current L2 of Polish Romani, and an older contact language for the other dialects of the group. Lithuanian, Latvian, and Estonian Romani have come under the influence of the Baltic languages Lithuanian and Latvian (e.g. Latvian Romani shows some Lithuanian loans as well as current influence of Latvian), and Estonian Romani shows some influence of its current contact language, Estonian. Russian is the current L2 of Russian Romani, and also the dominant current L2 of Lithuanian Romani. On the other hand, Russian influence is only sporadic in Latvian and Estonian Romani.

North Central

The North Central dialects are spoken in the Czech Republic, northern and central parts of Slovakia, southern Poland (the Podhalie region), western Ukraine (Ruthenia), and parts of Rumanian Transylvania. The original North Central dialects of the Czech Republic, viz. Bohemian and Moravian Romani, are now extinct due to extermination of their speakers. Instead, various North Central varieties from Slovakia have been spoken in the Czech Republic since the Second World War. Bohemian, Moravian and West Slovak Romani constitute the western subgroup of the North Central dialects, while East Slovak and South Polish Romani constitute the eastern subgroup. Varieties of Central Slovak Romani link the two subgroups. Gurvari is affiliated with the North Central dialects, although it now shows closer links to the North Vlax Lovari due to interdialectal contact. We use names of regions and localities to identify individual varieties of Slovak Romani more precisely (e.g. the Lučivná, Zborov, and Humenné varieties of East Slovak Romani, and the Balog and Pribylina varieties of Central Slovak Romani). All North Central dialects share Serbian/Croatian and Hungarian as older L2s. The current L2s are Slovak, Czech, Polish or, rarely, Hungarian.

Table I.4. North Central dialects of Romani

Dialect	Location	Source
Bohemian R	W Czech Republic	Puchmayer 1821
West Slovak R	W Slovakia	von Sowa 1887
Central Slovak R (various)	C Slovakia	RMS (Support Elicitation)
East Slovak R: Humenné	E Slovakia	Lípa 1963
East Slovak R (general)	E Slovakia	Hübschmannová, Šebková, and Žigová 1991
East Slovak R (various)	E Slovakia	RMS (Support Elicitation)
South Polish R	S Poland	RMS (Support Elicitation)
Gurvari	Hungary	Vekerdi 1971

South Central

The South Central dialects are spoken in southern Slovakia, northern and western Hungary, eastern Austria (the Burgenland region), and northeastern Slovenia (the Prekmurje region). Significant outmigrant communities from Slovakia live in the Czech Republic (Rumungro). Speakers of the South (as well as the North) Central dialects are frequently referred to as *Rumungri* by other Rom groups. Linguists, however, tend to use the term *Rumungro* for the northern subgroup of the South Central dialects, as spoken in southern Slovakia and northern Hungary. We sometimes use the term *older Rumungro* to refer to the idiolect of János Sípos as recorded by Müller (1869). The South Central dialects of western Hungary (Vend), Austria (Roman), and Slovenia (Prekmurje) form the southern, or Vendic, subgroup. Hungarian is the most significant (current or recent) L2 of all South Central dialects, including those of Prekmurje. The South Central dialects also share older influences from Croatian (or, more precisely, Ikavic Serbian/Croatian). Slovak is the current L2 of Klenovec Rumungro, German of Roman, and Slovene of Prekmurje. The current L2 of the other dialects is Hungarian.

Table I.5. South Central dialects of Romani

Dialect	Location	Source
Rumungro: Klenovec	S Slovakia	RMS (Support Elicitation)
Rumungro: Šóka	SW Slovakia	Elšík, in preparation
Rumungro: Sípos	N Hungary or S Slovakia	Müller 1869
Rumungro: Nógrád	N Hungary	RMS (Support Elicitation)
Vend	W Hungary	Vekerdi 1984
Roman	E Austria	Halwachs 1998

Table I.6. Slovene and Apennine dialects of Romani

Dialect	Location	Source
Slovene R	E Slovenia	RMS (Cech and Heinschink)
Abruzzian R	S Italy	Spinelli (ms.)
Calabrian R	S Italy	Soravia 1977

Slovene and Apennine

Slovene Romani (referred to by a plethora of terms in Romani linguistics, such as *Harvati* 'Croatians' or *Sinti Istriani*) is spoken in Slovenia, Istria, and northeastern Italy. The Apennine dialects are spoken in southern Italy, in the regions of Abruzzi, Molise, and Calabria. Both groups show important similarities to the Arli-type South Balkan dialects (see below), although they are both significantly distinct to be recognised as separate dialect groups. Slovene Romani is geographically and linguistically transitional to the South Central and the Sinti dialects, and the Apennine dialects share certain developments with Italian Sinti. Our sample includes a variety of Slovene Romani as spoken in the Dolenjski region of Slovenia, and Abruzzian and Calabrian Romani of the Apennine group. The recent contact L2 of Slovene Romani was Croatian, while the current L2 is Slovene. Italian varieties of Slovene Romani are in current contact with Italian. Abruzzian Romani and other Apennine dialects show significant influences from south Italian dialects.

South Balkan

The South Balkan dialects (originally referred to as *South Balkan I* in Romani linguistics, Boretzky 1999a) form a very heterogenous group within the Balkan dialects. They are spoken in the countries of the southern Balkans (viz. Serbia, Kosovo, Macedonia, Albania, Greece, and Bulgaria) as well as in Turkey, Iran, Rumania, Ukraine, Russia, and Georgia. Significant communities of recent outmigrants from Serbia and Macedonia live in Austria, Germany and other western European countries. *Arli-type* dialects are spoken in Serbia, Kosovo, Macedonia, and northern Greece. Among them, southern Arli varieties of southern Macedonia and Greece (Prilep, Florina and Karditsa) appear to form one subgroup, and northern Arli varieties of Serbia, Kosovo and northern Macedonia (Gilan, Kumanovo, Skopje) appear to form another subgroup. The dialect of Prizren is distinct from both. We refer to Zargari and Romano as *Iranian* Romani. The variety of Rumelian Romani in our sample is the one named "Sedentary" in Paspati (1870).

The current L2s are those of the respective countries (Albanian alongside Serbian in the case of Prizren and Gilan Arli, Russian in the case of Crimean Romani, Azeri alongside Persian in the case of Iranian Romani). Arli of Florina and Karditsa shows recent Macedonian influences, Epiros Romani has had some contact with Albanian, Sepečides with South East Slavic and Greek, and Prilep Arli and Varna Bugurdži with Turkish. Crimean Tatar is the recent L2 of Crimean Romani.

Table I.7. South Balkan dialects of Romani

Dialect	Location	Source
Arli: Prizren	S Serbia (Kosovo)	RMS (Igla)
Arli: Gilan	S Serbia (Kosovo)	Boretzky 1996a
Arli: Kumanovo	N Macedonia	RMS (Support Elicitation)
Arli: Skopje (Baručisko)	N Macedonia	Boretzky 1996a
Arli: Prilep	S Macedonia	Boretzky 1997, RMS (Cech and Heinschink)
Arli: Florina	N Greece	RMS (Sechidou)
Arli: Karditsa	N Greece	RMS (Support Elicitation)
Epiros R	NW Greece	RMS (Support Elicitation)
Sepečides	W Turkey	Cech and Heinschink 1999
(Sofia) Erli (older)	W Bulgaria	Boretzky 1998
(Sofia) Erli (modern)	W Bulgaria	RMS (Support Elicitation)
Yerli: Velingrad	SW Bulgaria	RMS (Support Elicitation)
Yerli: Rakitovo	SW Bulgaria	RMS (Support Elicitation)
Rumelian R	NW Turkey	Paspati 1870
Varna Bugurdži	NE Bulgaria	RMS (Support Elicitation)
Crimean R	S Ukraine, S Russia	RMS (Support Elicitation)
Zargari	N Iran	Windfuhr 1970; Baghbidi 2003
Romano	N Iran	Djonedi 1996

Balkan *zis*-dialects

The Balkan *zis*-dialects, named after their characteristic form for the word 'day' (< *dives*), have also been referred to as *South Balkan II* or the Bugurdži-Kalajdži-Drindari group (cf. Boretzky 2000). They form a relatively homogenous group within the Balkan dialects. They are spoken troughout Bulgaria, especially in the east and in the north of the country, and also in Macedonia and Kosovo (Bugurdži) and in southern Rumania (Spoitori). The dialects of the *Drindari* of Razgrad, the *Drindari* of Šumen, and the *Muzikanta* 'Musicians' (also called *Kutkadžus*, i.e. those who say *kutka* for 'there') of Sliven may be subsumed under the label *Drindari-type* dialects. Although an alternative name for the *Gadžikano* 'Non-Rom' dialect of Varna is *Drindari*, too, the dialect is more closely related to the dialect of Kaspičan. The dialect of *Nange Roma* 'Poor Rom' of Sliven is distinct from that of *Muzikanta*, which is spoken in a different neighbourhood of the town. An alternative name for the *Malokonare* is *Kalajdži* 'Tinner', and the dialect is closely related to that of the *Kalajdži* of Montana. An alternative name for the *Bugurdži* 'Drillmaker' dialect is *Kovački* 'Blacksmith'. Bulgarian is the current L2 of all *zis*-dialects spoken in Bulgaria. While they all share older Turkish influences, the Gadžikano and Kaspičan dialects still retain Turkish, alongside Bulgarian, as their current L2. The current L2 of Kosovo Bugurdži is Serbian, while only some speakers also know Albanian.

Table I.8. Balkan *zis*-dialects of Romani

Dialect	Location	Source
(Pazardžik) Malokonare	SW Bulgaria	RMS (Support Elicitation)
Montana Kalajdži	NW Bulgaria	RMS (Support Elicitation)
Kaspičan	NE Bulgaria	RMS (Support Elicitation)
(Varna) Gadžikano	NE Bulgaria	RMS (Support Elicitation)
Razgrad Drindari	NE Bulgaria	Kenrick 1967
Šumen Drindari	NE Bulgaria	RMS (Support Elicitation)
(Sliven) Muzikanta	E Bulgaria	RMS (Support Elicitation)
(Sliven) Nange	E Bulgaria	RMS (Support Elicitation)
Kosovo Bugurdži	Serbia (Kosovo)	Boretzky 1993a

North Vlax

The North Vlax dialects form a homogenous subgroup within the Vlax dialects. Great numbers of speakers migrated out of Transylvania (*Lovari* 'Horsedealers', *Cerhari* 'Tent-dwellers', and related groups) and Wallachia (*Kalderaš* or *Kelderar* 'Kettlemakers' and related groups) during the nineteenth century into most European countries as well as overseas, especially into the Americas. Nevertheless, North Vlax dialects are still widely spoken in all parts of Rumania and Moldavia (the *Rakarengo* dialect is spoken by a clan that migrated from Moldavia to Rumania some fifty years ago). Some of the dialect names (e.g. *Bougeští*, *Taikon*, *Markuleš*) derive from clan names. All North Vlax dialects share significant older Rumanian influence, and Lovari var-

Table I.9. North Vlax dialects of Romani

Dialect	Location	Source
Cerhari	Hungary	Mészáros 1976
Rakarengo	Moldavia/Rumania	RMS (Support Elicitation)
Lovari: Austrian	Austria	Cech and Heinschink 1998
Lovari: Hungarian	Hungary	Hutterer and Mészáros 1967
Lovari: Polish	Poland	Pobożniak 1964
Lovari: Bougešťi	Czech Republic, Slovakia	RMS (Elšík)
Lovari: Norwegian	Norway	Gjerde 1994
Lovari/Kelderaš	Germany	Matras 1994a
Kalderaš: Taikon	Sweden	Gjerdman and Ljungberg 1963
Kalderaš: Bunkuleš	Serbia	Boretzky 1994
Kalderaš: Markuleš	Serbia (Vojvodina)	Boretzky 1994
Kalderaš: Italian	Italy	Soravia and Fochi 1995

ieties and Cerhari in addition share more recent Hungarian influence. The recent and current L2s of individual varieties reflect the migration route and the current location of individual groups. Thus, for example, Bougešťi of the Czech Republic shows Slovak and Czech contact layers, Norwegian Lovari has been in contact with French and Norwegian, Taikon Kalderaš has borrowed from Russian and Swedish, and Italian Kalderaš shows Serbian and Italian influences.

South Vlax

The South Vlax dialects are spoken in Croatia, Bosnia, Serbia and Montenegro, Albania, Macedonia, Bulgaria, Greece, Rumania (Wallachia), and Turkey, and by older outmigrant communities in Italy. Significant communities of recent outmigrants from Serbia and Macedonia live in Austria, Germany and other western European countries. The Gurbet dialects spoken in Serbia, Kosovo, and Macedonia (also called *Džambazi* in Macedonia), Dasikano of Montenegro, and Xoraxane of Italy constitute the *Gurbet-type* dialects. An alternative name for the Rešitare is *Čergare*. All South Vlax dialects share an older Rumanian influence (prior to their migration out of Wallachia, which certainly took place long before the outmigration of the North Vlax speakers, cf. Boretzky 2003). The current L2 of the dialects of Bosnia, Serbia and Montenegro is Serbian, supplemented by Albanian in case of Priština Gurbet. Macedonian, Bulgarian, and Greek are the current L2s of the dialects of Macedonia, Bulgaria, and Greece, respectively. Turkish was an important L2 for Ajia Varvara, as it still is for some of the South Vlax dialects of Bulgaria (Rešitare, Varna Kalajdži, Vălči Dol, and Kalburdžu).

Table I.10. South Vlax dialects of Romani

Dialect	Location	Source
Gurbet: Priština	Serbia (Kosovo)	Boretzky 1986
Gurbet: Kumanovo	Macedonia	RMS (Support Elicitation)
Gurbet: Bačka	Serbia (Vojvodina)	RMS (Support Elicitation)
Dasikano	Montenegro	RMS (Boretzky)
Xoraxane	Italy	Franzese 1986
Ajia Varvara	Greece	Igla 1996
Rešitare	SW Bulgaria	RMS (Support Elicitation)
Varna Kalajdži	NE Bulgaria	RMS (Support Elicitation)
Vălči Dol	NE Bulgaria	RMS (Support Elicitation)
(Sindel) Kalburdžu	NE Bulgaria	RMS (Support Elicitation)
(Vidin) Cocomanya	NW Bulgaria	RMS (Support Elicitation)
Lom	NW Bulgaria	RMS (Support Elicitation)

Ukrainian

Ukrainian Romani is spoken by the long-settled Rom of eastern Ukraine and adjacent areas of southwest Russia. The Ukrainian Romani dialects are closely affiliated with the Vlax dialects, although they do not exhibit some of the defining features of prototypical Vlax. The current L2 of Ukrainian Romani is Ukrainian and, to a lesser extent (at the time of documentation), also Russian. Rumanian is an older L2.

Table I.11. Ukrainian dialects of Romani

Dialect	Location	Source
Ukrainian R	Ukraine	Barannikov 1934

Alphabetical list of dialects, their locations and map index

Table I.12. Alphabetical list of Romani dialects

Dialect	Index no. on map	Location
Abruzzian R	24	S Italy
Ajia Varvara	26	Greece
Arli: Florina	3	N Greece
Arli: Gilan	9	S Serbia (Kosovo)
Arli: Karditsa	1	C Greece
Arli: Kumanovo	7	N Macedonia
Arli: Prilep	5	S Macedonia
Arli: Prizren	8	S Serbia (Kosovo)
Arli: Skopje (Baručisko)	6	N Macedonia
Bohemian R	45	W Czech Rep.
Calabrian R	25	S Italy
Central Slovak R	47	C Slovakia
Cerhari	50	Hungary
Crimean R	27	S Ukraine, S Russia
Dasikano	28	Montenegro
East Slovak R	48	E Slovakia
English R	68	Great Britain (England)
Epiros R	2	NW Greece
Estonian R	64	Estonia
Finnish R	53	Finland
Gurbet: Bačka	29	Serbia (Vojvodina)
Gurbet: Kumanovo	7	Macedonia
Gurbet: Priština	30	Serbia (Kosovo)

Gurvari	51	Hungary
Kalderaš: Bunkuleš	31	Serbia
Kalderaš: Italian	(32)	Italy
Kalderaš: Markuleš	33	Serbia (Vojvodina)
Kalderaš: Taikon	70	Sweden
Kaspičan	20	NE Bulgaria
Kosovo Bugurdži	30	Serbia (Kosovo)
Latvian (Curland) R	63	Latvia
Lithuanian R	62	Lithuania
Lom	21	NW Bulgaria
Lovari/Kelderaš	71	Germany
Lovari: Austrian	34	Austria
Lovari: Bougešťi	35	Czech Rep., Slovakia
Lovari: Hungarian	36	Hungary
Lovari: Norwegian	72	Norway
Lovari: Polish	73	Poland
Montana Kalajdži	15	NW Bulgaria
Pazardžik Malokonare	13	SW Bulgaria
Podolie R	67	NW Ukraine
Polish R	61	Poland
Rakarengo	44	Moldavia/Rumania
Razgrad Drindari	16	NE Bulgaria
Rešitare	11	SW Bulgaria
Roman	37	E Austria
Romano		N Iran
Rumelian R	12	NW Turkey
Rumungro: Klenovec	38	S Slovakia
Rumungro: Nógrád	39	N Hungary
Rumungro: Sípos	40	N Hungary or S Slovakia
Rumungro: Šóka	41	SW Slovakia
Russian R	65	Russia
Sepečides	4	N Greece
Šumen Drindari	17	NE Bulgaria
Sindel Kalburdžu	18	NE Bulgaria
Sinti: Austrian	56	Austria
Sinti: German	54	Germany
Sinti: Hungarian	57	Hungary
Sinti: Lalere	55	Germany, Czech Rep.
Sinti: Lombardian	60	N Italy
Sinti: Manuš	58	France
Sinti: Piedmontese	59	N Italy, S France
Sliven Muzikanta	22	E Bulgaria
Sliven Nange	22	E Bulgaria

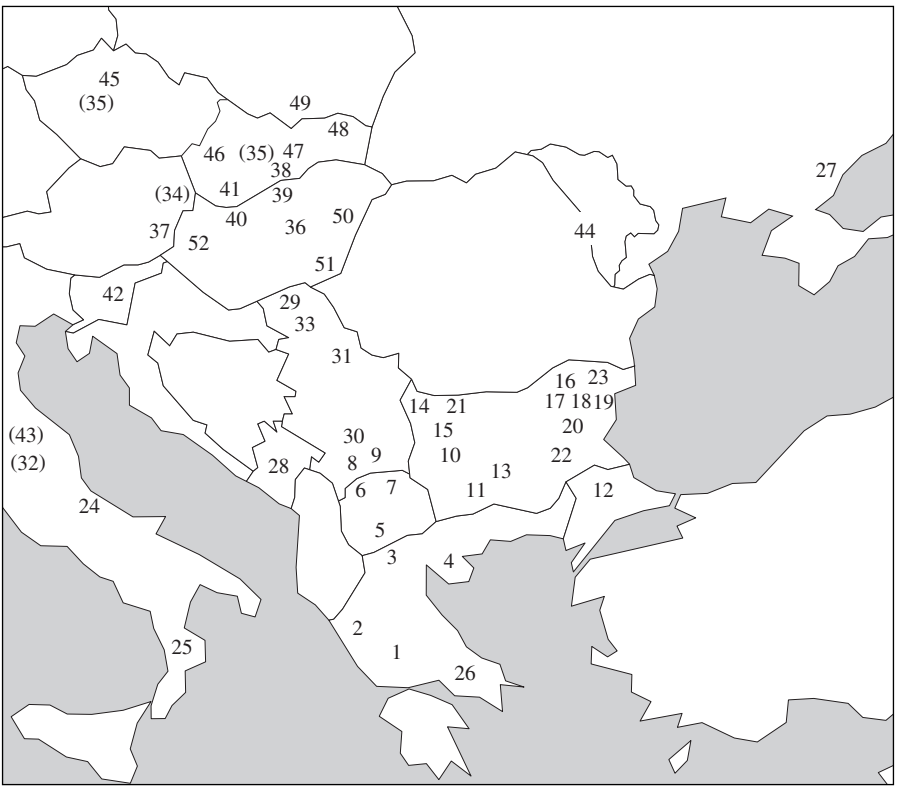
Table I.12. (*cont.*)

Dialect	Index no. on map	Location
Slovene R	42	E Slovenia
Sofia Erli	10	W Bulgaria
South Polish R	49	S Poland
Ukrainian R	66	Ukraine
Vălči Dol	23	NE Bulgaria
Varna Bugurdži	19	NE Bulgaria
Varna Gadžikano	19	NE Bulgaria
Varna Kalajdži	19	NE Bulgaria
Vend	52	W Hungary
Vidin Cocomanya	14	NW Bulgaria
Welsh R	69	Great Britain (Wales)
West Slovak R	46	W Slovakia
Xoraxane	43	Italy
Yerli: Rakitovo	11	SW Bulgaria
Yerli: Velingrad	11	SW Bulgaria
Zargari		N Iran

Table I.13. List of Romani dialects by map index number

No.	Dialect	Location
	Romano	N Iran
	Zargari	N Iran
1	Arli: Karditsa	C Greece
2	Epiros R	NW Greece
3	Arli: Florina	N Greece
4	Sepečides	N Greece
5	Arli: Prilep	S Macedonia
6	Arli: Skopje (Baručisko)	N Macedonia
7	Arli: Kumanovo	N Macedonia
7	Gurbet: Kumanovo	Macedonia
8	Arli: Prizren	S Serbia (Kosovo)
9	Arli: Gilan	S Serbia (Kosovo)
10	Sofia Erli	W Bulgaria
11	Rešitare	SW Bulgaria
11	Yerli: Rakitovo	SW Bulgaria
11	Yerli: Velingrad	SW Bulgaria

12	Rumelian R	NW Turkey
13	Pazardžik Malokonare	SW Bulgaria
14	Vidin Cocomanya	NW Bulgaria
15	Montana Kalajdži	NW Bulgaria
16	Razgrad Drindari	NE Bulgaria
17	Šumen Drindari	NE Bulgaria
18	Sindel Kalburdžu	NE Bulgaria
19	Varna Bugurdži	NE Bulgaria
19	Varna Gadžikano	NE Bulgaria
19	Varna Kalajdži	NE Bulgaria
20	Kaspičan	NE Bulgaria
21	Lom	NW Bulgaria
22	Sliven Muzikanta	E Bulgaria
22	Sliven Nange	E Bulgaria
23	Vălči Dol	NE Bulgaria



Map 1. Locations of Romani dialects in southeastern and central Europe

Table I.13. (*cont.*)

No.	Dialect	Location
24	Abruzzian R	S Italy
25	Calabrian R	S Italy
26	Ajia Varvara	Greece
27	Crimean R	S Ukraine, S Russia
28	Dasikano	Montenegro
29	Gurbet: Bačka	Serbia (Vojvodina)
30	Gurbet: Priština	Serbia (Kosovo)
30	Kosovo Bugurdži	Serbia (Kosovo)
31	Kalderaš: Bunkuleš	Serbia
32	Kalderaš: Italian	Italy
33	Kalderaš: Markuleš	Serbia (Vojvodina)
34	Lovari: Austrian	Austria
35	Lovari: Bougešťi	Czech Rep., Slovakia
36	Lovari: Hungarian	Hungary
37	Roman	E Austria
38	Rumungro: Klenovec	S Slovakia
39	Rumungro: Nógrád	N Hungary
40	Rumungro: Sípos	N Hungary/S Slovakia
41	Rumungro: Šóka	SW Slovakia
42	Slovene R	E Slovenia
43	Xoraxane	Italy
44	Rakarengo	Moldavia/Rumania
45	Bohemian R	W Czech Rep.
46	West Slovak R	W Slovakia
47	Central Slovak R	C Slovakia
48	East Slovak R	E Slovakia
49	South Polish R	S Poland
50	Cerhari	Hungary
51	Gurvari	Hungary
52	Vend	W Hungary
53	Finnish R	Finland
54	Sinti: German	Germany
55	Sinti: Lalere	Germany, Czech Rep.
56	Sinti: Austrian	Austria
57	Sinti: Hungarian	Hungary
58	Sinti: Manuš	France
59	Sinti: Piedmontese	N Italy, S France
60	Sinti: Lombardian	N Italy
61	Polish R	Poland

Map 2. Location of Romani dialects outside southeastern and central Europe

62	Lithuanian R	Lithuania
63	Latvian (Curland) R	Latvia
64	Estonian R	Estonia
65	Russian R	Russia
66	Ukrainian R	Ukraine
67	Podolie R	NW Ukraine
68	English R	Great Britain (England)
69	Welsh R	Great Britain (Wales)
70	Kalderaš: Taikon	Sweden
71	Lovari/Kelderaš	Germany
72	Lovari: Norwegian	Norway
73	Lovari: Polish	Poland

Notes

Chapter 1

No notes

Chapter 2

1. Givón (1990) uses the term *frequency distribution*.
2. Also mentioned as relevant for cognitive complexity is order of acquisition (Givón 1990: 953).

Chapter 3

No notes

Chapter 4

1. Romani is also spoken in emigrant communities originating from Europe, e.g. in the Americas, but also in Azerbaijan (see Appendix I: Sample dialects).
2. As documented for the central or inner languages of India, where the ancestor language of Romani is believed to have emerged.
3. In a number of recent publications, Hancock (e.g. 1998, 2000) claims that Romani was formed as a military koiné by a caste of warriors assembled to resist the Islamic invasions of India. In some circles, this view is gaining popularity as it claims to revise what is referred to as potentially racist or at least stereotypical images of the Rom. There is, however, neither linguistic nor historical evidence to support it.
4. For studies of the emergence of institutionalised norms in some Romani-speaking communities see Hübschmannová (1995), Friedman (1999), Halwachs (1996), as well as Matras (1999e).
5. Compiled by Yaron Matras, Viktor Elšík, Katrin Hiietam, Christa Schubert, Barbara Schrammel, and Irene Sechidou; University of Manchester, November 2001.
6. To date (December 2003), questionnaire elicitations have been carried out in Bulgaria, Rumania, Greece, Macedonia, Serbia, Hungary, Slovakia, Czech Republic, Poland, Estonia, Lithuania, and Finland.

Chapter 5

1. The following graphemes in our transcription of Romani are distinct from IPA symbols: the alveolars *c* [ʦ] *dz* [ʣ], the postalveolars *š* [ʃ] *ž* [ʒ] *č* [ʧ] *dž* [ʤ], the alveo-

palatals *ś* [ɕ] *ź* [ʑ] *ć* [tɕ] *dź* [dʑ], the palatals *ṭ'* [c] *ḍ'* [ɟ] *ň* [ɲ] *ľ* [ʎ], the velarised lateral *ł* [lʸ], and the high central unrounded vowel *y* [ɨ]. We use subscript dots for retroflexes, *h* for aspiration, apostrophe for palatalisation, macron for vowel length: e.g. *ṭ* [ʈ], *ṭ'* [tʲ], *th* [tʰ], *ā* [a:]. The grapheme *ř* is etymological, standing for various rhotic sounds, depending on dialect: e.g. uvular [ʁ], geminate [r:] etc. The grapheme *x* stands for the uvular [χ] or the velar [x], depending on dialect.

Chapter 6

1. In a few dialects in contact with Hungarian, borrowed personal names may be markerless in the vocative singular rather than in the nominative singular (e.g. *Alen* from the female name *Alen-a*, or *Bēluš* from the male name *Bēluš-i* in Šóka Rumungro). The markerless vocatives are direct loans of Hungarian markerless or truncated vocatives.
2. The final /*s*/ of the perfective person–number marker **-as* has been lost in the preterite but not in the pluperfect. Note that the loss of /*s*/ in word-final positions is not a general phonological process in the Northeastern dialects.
3. The third-person pluperfect forms have been formed anew from the corresponding preterite forms, and do not continue the Early Romani forms (cf. the expected 3SG **-ah-ahi* and 3PL **-e-h-ahi*).
4. Non-perfective and perfective forms are represented by person–number markers as they occur, respectively, in the so-called short (subjunctive) forms and in the preterite.
5. The preterite–pluperfect homonymy in the third-person singular and plural in Finnish and Rumelian Romani (see Chapter 13) is not indicative of any number hierarchy, and will not be discussed here.
6. There is evidence from other Sinti dialects that the present set arose through erosion of the so-called long forms in *-a*. This is still visible in Hungarian Sinti in the second-person singular present suffix *-ē* (< **-eh* < **-eh-e* < **-es-a*). We may assume parallel developments for the first- and third-person singular present suffixes: *-av* (< **-av-a*) and *-el* (< **-el-a*).
7. Thus, the stress is initial in Central dialects influenced by Hungarian, West Slovak, and Czech, and it is penultimate in dialects influenced by East Slovak.
8. Masculine-to-feminine extensions with demonstrative and the article are attested in a few other dialects, too, but there they only concerns certain variants, and do not lead to gender neutralisation (see Chapter 8).
9. The third-person singular and plural pluperfect forms are also homonymous in Austrian and Slovak Lovari (e.g. *kerd-oun* 'they would have done'). However, the origin of the suffix *-oun* is not obvious, and so we are not in a position to evaluate the direction of the extension.
10. Only a few dialects have created plural forms of the personal interrogative: cf. the plural oblique stem *kon-en-* in Austrian Sinti, Welsh Romani, and Finnish Romani. A nominative plural form is only attested in Finnish Romani (*kon-e*).

Chapter 7

1. We disregard initial vowels of the roots (e.g. *is-*, *es-*, and *ih-*, *eh-*), as they are only present in some dialects, often optionally. We also disregard obvious secondary modifications of the two roots: *š-* in Welsh and Epiros Romani instantiates the root *s-*, and a zero root in Piedmontese Sinti, and variantly also in Arli of Prizren, Gilan, and Skopje, instantiates the root *h-*.
2. In a few dialects, there are similar erosion processes in the inflection of non-middle verbs with a stem ending in */Vv/* (e.g. Roman *dar-ava-v* 'I frighten', *dar-aj-s* < **dar-ave-s* 'you.SG frighten', *dar-a-l* < **dar-ave-l* 's/he frightens'). As the erosion in these verbs follows the same person hierarchy as the erosion in middle verbs, we have decided to drop the discussion.
3. Certain idiosyncracies of Finnish Romani, such as the one just mentioned, may be motivated by the special sociolinguistic status of the dialect (see Appendix I: Sample dialects).
4. In some dialects the second-person plural has been completely taken over by the third-person plural (see Section 7.4), and so there is no number homonymy in the second person any more either.
5. Austrian Lovari shows the same pattern in the third person (i.e. two pluperfect subsets and number homonymy) as the closely related Bougešti, but there is also a number homonymy in the second person of the pluperfect (see above). This means that the third-person homonymy may be viewed as licenced by the second-person homonymy.
6. There is still some vacillation in the use of the forms in our corpus, but the tendency to employ the forms with the remoteness marker for the plural and not for the singular is reasonably strong.
7. The precise distribution of these sets differs from dialect to dialect. In some dialects, reduced accusative forms may develop into enclitics (e.g. Prilep Arli *oles dik-lum-les* 'it was him who I saw'), and even into suffixes (e.g. Epiros *olesko dad bišavel-əs ti dajate* 'his father sends him to (his) mother'; but cf. Chapter 5 for an alternative scenario).
8. We suggest that the commonality of this development partly serves the need to avoid number homonymy between the second-person plural and the second-person singular forms that had been present in many dialects.
9. With some likelihood, the person-neutralising extension was triggered, or at least allowed, by contact with Ukrainian and/or Russian. East Slavic languages (unlike West and South Slavic) do not encode person in past verb forms. Person deixis is recovered from free personal pronouns (e.g. Ukrainian *ja skazav* 'I.M said' – *ty skazav* 'you.SG.M said' – *vin skazav* 'he said').
10. The perfective first-person plural suffix has an identical shape (viz. *-em*) also in Yerli and Malokonare. In these dialects, however, the development is purely phonological (**ja* > *e*): cf. Velingrad **kerdjam* > *kərdem* 'we did', but also **kerdjas* > *kərdes* 's/he did', **phenja* > *phəne* 'sisters', or **tasja* > *tase* 'tomorrow'.

11. Here and in the following forms, /*V*/ stands for an abstract vowel, which acquires its quality through vowel harmony with adjectival inflections.
12. In all examples in this subsection, the matrix verb and the complement verb are underlined; only verb forms are fully segmented.
13. In many dialects (e.g. in Bohemian, Slovene and Ukrainian Romani, Lovari, and Sinti), the new infinitive coexists with finite modal complements.
14. In our corpus, the variation does not seem to correlate with any structural or semantic property of the complementation construction (such as the person or number of the matrix verb, the semantics of the matrix or the complement verbs, or the presence/absence of the complementiser).
15. Unlike personal pronouns, the reflexive pronouns have no nominative forms in Romani, due to syntactic constraints.
16. At the same time, the pattern is an instance of sound iconicism and – assuming that certain phonemes are in some sense more complex, or "stronger", than others (e.g. /*a*/ being more sonorous than /*e*/) – we could add that the first person is more complex than the second and third person.
17. There is no pluperfect in Slovene Romani.

Chapter 8

1. This masculine nominative singular form (*ova*) is attested in the closely related dialects of northwestern Slovakia.
2. The same development is found in Domari, while all other Indo-Aryan languages display gender distinction(s) in the plural of declinable adjectives.

Chapter 9

1. This differentiation asymmetry has been transferred from German, together with the borrowing of the adverbial superlative marker *am*.

Chapter 10

1. In most dialects, the negative past form is a negated past form (e.g. Northeastern *sys* vs. *na-sys*, Slovene Romani *hine* vs. *na-hine*, Core Sinti *his* vs. *his nit*).
2. It seems more likely though, that vowel-initial forms of the ability particle *ašti* are original and that the inability particle was created as the former's regular negation: *našti* < **na-ašti*. See also Section 10.4.
3. Only in a few dialects is there no such clear complexity asymmetry in indigenous modal particles: e.g. Lithuanian Romani *š-ašty* 'can' vs. *n-ašty* 'cannot', Bunkuleš Kalderaš *d-ašti-* vs. *n-ašti*.
4. The cross-dialectal distribution of this phenomenon is due to convergence and/or borrowing from contact languages.

5. Some dialects, especially the Vlax dialects, also show distinct negators for lexical verbs and the copula (e.g. Bunkuleš Kalderaš *či* vs. *naj*). This distinction in negators is always reflected in the inflection of verbs.
6. The particle **ni* is rarely attested as an emphatic enclitic: e.g. Slovene Romani *me-ni* 'I', Sofia Erli *oj-da-ni* 'he too', Zargari *xā-ni* 'I eat' (cf. Boretzky 1995: 25–26, Elšík 2000c). It is distinct both from the negator *ni* 'not' and from the negative focus particle *ni* 'either; not even', which is a loan from South Slavic.
7. Although copula forms usually grammaticalise into necessity modals (e.g. *si-te*, *hun-te* 'must').
8. Borrowing of the YES-particle is also attested in other languages of the area (and elsewhere). For example, South/East Slavic *da* is used in Rumanian, and German *ja* (*jo*) in colloquial Czech and dialectal Slovene.

Chapter 11

1. Greek-derived cardinals (see Section 11.5) do not inflect in modifier position. The lack of inflection here may be connected to the fact that the Greek numerals all end in a vowel and do not easily take the vocalic oblique suffix (e.g. **efta-e* '7'). We have little comparative data on the inflection of compound numerals. However, at least in some dialects they may inflect through the inflection of the last element (e.g. West Slovak Romani *biš-the-štār-e* '24' or Russian Romani *trin-deš-e* '30').
2. Either due to dialect-specific erosion, or due to contamination by the Greek equivalent *tritos* '3rd' already in Early Romani. In the latter case, the from *trin-to* is a secondary regularisation.
3. It is attested as a clausal (but not constituent) coordinator, for example, in some varieties of Slovak Romani.
4. Bakker (2001) adduces similar examples.
5. Only two dialects of our sample, Sípos Rumungro and Roman, are attested as making use of the connector *-u-* with compound tens: *eftavārdeš-u-efta* '77'.
6. Kalburdžu represents a deviant type in that the zero connector is used with ten+units based on '10' and '30', but not in the intervening ten+units based on '20', where *-taj-* is employed.
7. A grammaticalisation of the conjunctive connector *u* 'and', before it came to be restricted to clausal coordination.
8. A few dialects in Bakker's (2001: 100) sample have the multiplicative *deš deša* '100' (= '10 10-PL').
9. At least in some of these dialects (e.g. in Rumungro), there is also *biš-šel* '2,000' (= '20–100'), *šōvārdeš-šel* '6,000' (= '60–100') etc.
10. However, an alternative Indic etymology has also been proposed (Mānušs, Neilands, and Rudevičs 1997).
11. In these dialects, the cognitive focus in constructing, for example, the numeral '60' (= the sixth ten of the first hundred) is on 'hundred' rather than 'ten'. A similar

system is Gothic *sibuntéhund* '70', *ahtautéhund* '80', *niuntéhund* '90'. We have no data on how '600' is expressed in the relevant Balkan dialects.
12. The numeral *oxto-nda* '80' is also attested in Rumelian Romani.
13. One exception is *štar-deset* '40', compounded of the indigenous *štar* '4' and the borrowed *deset* '10'. The retension of this anomaly is most likely due to the formal similarity to the Slovene numeral *štiri-deset*.
14. In some (perhaps most) dialects, the extent of the original Romani numerals is a sociolinguistic variable. For example in Slovak Romani of Bystrany, some speakers only know the original numerals up to '6', while others can count up to '39', although they rarely use these higher original numerals in natural discourse. The variation seems to be age related: the older know more Romani numerals than the younger. Note, however, that there is no language shift going on in the community. Numeral fusion may occur without language shift.
15. Note that the extensive borrowing of Greek tens in Karditsa Arli and Epiros, two dialects in current contact with Greek, is probably not a case of numeral fusion, since ten+unit numerals are internal compounds (e.g. Epiros *saranda-njekh* '41', with the indigenous unit numeral).
16. Unlike the loans for 'zero', 'million', 'milliard' etc., which are mostly fusional.
17. The Balkan Slavic numeral is itself a loan from Greek.
18. The Hungarian numeral is itself an old loan from Iranian.
19. The (dialectal) Slovene numeral is itself a loan from German.

Chapter 12

1. Note that the case marking of the antecedent of the subject of the embedded clause differs in the two examples. This, however, is of no importance to the contrast between subject-continuity and subject-discontinuity.

Chapter 13

1. However, we cannot posit zero-coding of the non-perfective with regard to the perfective, as it is not merely an absence of an overt marker that differentiates these values.
2. In German and Austrian Sinti, the secondary short forms are most common, while the long forms are rare, formal, variants. According to Valet (1991: 123), the secondary short forms are used in the non-past indicative of verbs that are followed by a clausal complement (e.g. *kamaw te kerap* 'I want to do'), while the long forms are used elsewhere (e.g. *kamava tut* 'I like you').
3. In fact, in many dialects, the remoteness suffix **-asi* appears to 'contain' the suffix *-a*. This might tempt the analyst to derive the imperfect forms from the corresponding long forms, and to assume a bimorphemic structure of the remoteness marker (i.e. **-a-si*). There are two arguments against such an analysis, a synchronic one and a

diachronic one. First, the remoteness suffix also occurs in the perfective where there are no forms in *-a* (cf. 1SG **kerdjom-asi* but no ***kerdjom-a*). Second, phonological change and resulting morphophonological alternations need not affect both suffixes in a parallel manner. For example, in Polish Romani, erosion affects the remoteness suffix (*-ys* < **-as*), but not the suffix *-a*. And in Rumungro, the two suffixes differ in alternations they trigger (cf. the */s/* > */h/* alternation in the long form *kereh-a* 'you.SG will do' but not in the non-perfective form *keres-ahi* 'you.SG were doing'). This means that, in language change, the remoteness marker is not treated as containing the suffix *-a-* of the long forms.

4. Most of the dialects that innovated remoteness marking by grammaticalisation of the copula (e.g. Arli of Skopje, Gilan, Prizren and, tentatively, Slovene Romani) would have lost the distinction between the original imperfect and the long forms through a regular deletion of the final */s/*. This suggests that the innovation was triggered by the formal conflation between the two sets of forms. Nevertheless, we find both dialects that permit such a conflation without innovating the distinction (see above), and dialects that have innovated remoteness marking without having undergone the conflation (e.g. Florina Arli, where the copula construction alternates with the original imperfect in *-as*; and Abruzzian Romani).
5. Modern Prilep Arli lacks the pluperfect altogether.
6. Even in dialects without the middle contractions, the presence of a reflex of the Early Romani non-perfective middle suffix **-jov-* is by itself a criterion of inflectional classification, as the suffix is restricted to a single class, viz. the middles. Although the non-perfective middle suffix is primarily a derivational marker, it also participates in verb inflection, in that it does not occur in perfective forms. In this sense, the non-perfective middle suffix forms part of inflections, rather than inflectional stems.
7. In some dialects the perfective stem of the verb *dža-* 'go' became more similar to the non-perfective stem by regular sound change (rather than morhological extension): e.g. Xoraxane-Dasikano-Gurbet and Kalburdžu *dže-l-* (< **ge-l-*), Ukrainian Romani *dži-l-* (< **gi-l-*), and Finnish Romani *jē-l-* < *džē-l-* (< **ge-l-*; cf. non-perfective *ja-* < *dža-*).
8. A similar outcome is found in Slovene Romani: the imperfect contains the 'strong' */s/* (e.g. *kers-e* 'you.SG were doing'), while the future contains the eroded */h/* (e.g. *kereh-a* 'you.SG will do'). However, in the case of Slovene Romani, hypotheses about the origin of the strong */s/* depend on assumptions about the origin of the non-perfective suffix *-e*.
9. In Ajia Varvara, the subjunctive negator *na* is also used in the future and with imperfect verb forms in conditionals (but not with other imperfect verb forms), i.e. in modal uses of indicative forms. It also optionally extends to the imperative.

Chapter 14

No notes

Chapter 15

1. Romani middles have been also referred to as *mediopassives* (cf. Bubeník and Hübschmannová 1998) or *synthetic passives* (cf. Boretzky and Igla 1994). The precise semantic domain of Romani middles varies from dialect to dialect. It commonly includes spontaneous events such as change of shape (e.g. 'stretch ITR'), physicochemical change ('burn ITR'), or disruption of object's material integrity (e.g. 'explode ITR'), but also positionals (e.g. 'be lying'), nontranslational motion (e.g. 'turn ITR'), and more. Middle passives compete with analytic passive constructions in Romani. Cf. Kemmer (1993) for a typological study of middle verbs.
2. A rare transitive middle is the verb *axal-jov-* 'to understand' in some dialects.
3. Haspelmath (1993) has shown that verb semantics (viz. the likelihood of spontaneous occurrence of an event) is a factor determining the direction of derivation, and thus morphological complexity, in pairs of causative and what he calles *inchoative* (= anticausative) verbs. We do not discuss the semantic distribution of Romani causative vs. anticausative derivations here. It can be noted, however, that causative morphology is more developed in Romani than in most of its European contact languages, which show preference for anticausative derivation (cf. Haspelmath 1993: 102). This Romani feature is an Indo-Aryan heritage, supported locally by the few European contact languages that do use morphological causatives more extensively, especially Turkish and Hungarian.
4. In the Central dialects, the historical valency-increasing suffix *-ker-* has developed into an aktionsart (iterative) marker. Derivations containing the 'complex' markers *-av-ker-*, *-ar-ker-*, *-ker-ker-* and similar are in fact iteratives of causatives, iteratives of factitives, or 'double' iteratives: cf. Šóka Rumungro *ker-* 'do, make' > *ker-av-* 'have sth. done' > *ker-av-ker-* 'have sth. done frequently'.
5. Various other morphological extensions that are found with valency-decreasing markers are also found with valency-increasing markers. In other words, if middle derivations contain a complex marker, then so do corresponding causatives and factitives: cf. Šóka Rumungro *cid-isaj-ov-* 'stretch ITR' *and cid-isaj-ār-* 'stretch TR' (< *cid-* 'pull').
6. When used as attributive modifiers, passive participles also encode case, in addition to number and gender. In predicative use, they are in the default, nominative, form.

Chapter 16

1. The plural forms in *-e* show nominative–oblique homonymy in the inflectional class of the example adjective, but not in other inflectional classes.
2. In other words, the accusative form of substantivals derives from the oblique stem by a zero derivation. This analysis, fully presented in Elšík (2000b), is based on Kibrik's (1991: 257) argumentation on the status of the ergative within noun paradigms in Daghestanian languages.

3. The distinction between local/temporal and other case roles is, of course, somewhat arbitrary. Nevertheless, since local/temporal relations represent a well-defined semantic domain that allows for a paradigmatic arrangement and, unlike other case roles, interacts with the category of orientation (see Chapter 18), we have decided to postulate a separate category of *localisation*. It is dealt with in detail in Chapter 17. The distinct character of local cases has been argued for in Kibrik (2003: 43).
4. We have little comparative data on some other case roles (such as agents in passive or causative constructions, quantitative difference etc.), and so we leave them out of discussion.
5. It was a particle rather than an adposition in Early Romani, as it still is outside of the Balkans, since it did not assign case to the following noun phrase (the standard of comparison or equation).
6. The palatalisation of velars before front vowels (e.g. *-g'e*, *-g'i* vs. *-go*) is phonotactic, i.e. automatic.
7. The deletion of a final /*s*/ is not restricted to the accusative, and hence rather phonotactic than morphological, in some other dialects.
8. Thus in Hungarian Lovari, the selection of the oblique inflection is determined not only by gender of the head noun (as in most dialects), but also by gender of the dependent genitive noun (unlike all other dialects).
9. Whenever there is inflectional assimilation of the xenoclitic class of adjectives to the oikoclitic class (e.g. in Welsh Romani, Core Sinti, and the South Central dialects), it occurs in all inflectional values, and thus does not contribute to differentiation asymmetries between the nominative and the oblique.
10. Complete inflectional assimilation in the oblique is very rare and appears to be restricted to reduced demonstratives (e.g. Lithuanian Romani *d-e* 'this' like *bar-e* 'big').
11. Western Central dialects have lexicalised the "sociative" *ratʰaha* of the noun 'night' to mean 'in the morning', via the reading 'with the passing night'.
12. The alternation ceased to be automatic in some dialects.
13. For example, the plural variants of the locative and ablative suffixes have undergone nasal assimilation in Finnish Romani (cf. singular *-es-te*, *-es-ta* vs. plural *-en-ne*, *-en-na*). Also, due to aspiration and further developments in the singular (**s* > *h* > *j* or *r*) and due to affrication and further developments in the plural (**s* > *c* > *č* > *dž*), dialect-specific number alternations were created in the sociative suffix (e.g. Sepečides *-sar* ~ *-džar*, South Polish Romani *-ja* ~ *-ca*), which had been uniform in Early Romani.
14. The copula agrees with the nominative Possessee in person and number and is marked for TAM categories.
15. There is some doubt about the Greek origin of the oblique intrusion *-on-* in xenoclitic adjectives.
16. A colloquial form corresponding to *-bAn* in Standard Hungarian.
17. In Vălči Dol, the adposition *ičin* may also be preposed, due to convergence with Bulgarian. Turkish, probably via Albanian, is also the source of the noun *sebepi*

'reason' in Kosovo Bugurdži, which may be used as a Reason case marker (e.g. *sebepi lengero* 'their reason' = 'because of them').
18. An identical form in Austrian and Hungarian Lovari might be due to diffusion from Rumungro. Šóka Rumungro shows the form *bisto*, a contamination of the loan *misto* and the indigenous preposition *bi*, which has the Substitutive role among its functions.
19. We consider the use of the Slovene comparative preposition *neg*(*o*) 'than' in the Exceptive function in Slovene Romani (e.g. *nikon naj hak aver neg amen* [nobody is.not here other than we] 'nobody else is here but us') to be periphrastic rather than a case of a proper Exceptive loan.
20. There are also dialects with loans in both functions (e.g. Austrian Sinti, Slovene Romani, Kumanovo Arli, Yerli, or Varna Bugurdži), and numerous dialects without loans in either function.
21. There are also dialects with loans in both functions (e.g. Austrian Sinti, Arli of Gilan, Kumanovo and Karditsa, Yerli, Varna Bugurdži), and numerous dialects with loans in neither.

Chapter 17

1. The term *localisation* is taken from Kibrik's (2003) study on inflectional case in Daghestanian languages. However, we extend the domain of the category from inflectional marking (cf. Kibrik 2003: 43–50) to local adpositions and adverbs as well.
2. The term *ground object* has been used by Talmy (1983).
3. Lack of spatial distance may be marked by proximate expressions or, more explicitely, by negated distant expressions (cf. the adverb *na-dur* 'close, near').
4. It should be noted that we use the terms *inessive*, *adessive* and *extraessive* as neutral with regard to orientation, in order to avoid the multiplicity and lesser transparency of traditional terms (such as *inessive* for stative inessive, *illative* for directive inessive, and *elative* for separative inessive).
5. There are also local adverbs derived from a few indigenous nouns denoting locations: e.g. *kher-e* 'at home, home' and *kher-al* 'from home' < *kher* 'house' (and similar derivations from nouns meaning 'place', 'side', 'road', 'village', and more). These local adverbs are more lexical than the local adverbs given in Table 17.3, most of which are synchronically underived, medial adverbs being an exception (cf. the noun *maškar* 'middle' in some dialects). We do not consider the concrete semantics of the lexical place adverbs to constitute localisation values.
6. In Polish Romani, due to complete replacement of the locative by the ablative, the latter is also used in non-separative localisations.
7. The (separative) ablative is more common than the (non-separative) locative in the core localisations. This asymmetry, however, is irrelevant for the category of localisation.
8. The absence of ablative encoding in some of the three core localisations in numerous dialects (e.g. Welsh and Bohemian Romani, Roman, Arli of Gilan, Prilep and Flo-

rina, Sepečides, Muzikanta, Kosovo Bugurdži, Taikon Kalderaš, and Ajia Varvara) may be due to gaps in our data.
9. See also Table 17.5 for Šóka Rumungro paradigm, which conforms to the hierarchy.
10. However, the complex adposition may also contain the specifically separative adposition *kat* 'from' in the separative reading (e.g. *angla kat man-dar* 'from the front of me').
11. A complete replacement by the adessive adpositions *te* 'at, to' and *kat* (< **katar*) 'from' had probably taken place in Florina Arli, too. Here, however, a specific separative inessive adposition is being secondarily grammaticalised from the corresponding local adverb: cf. *an*(*g*)*ral* 'out of' < *angral* 'from inside'.
12. It appears that the non-separative inessive example in (21b) in fact represents an extended use of the separative inessive *pal*. A "deep" separative, which derives from the reading 'I have been coming home from many countries' (inferred from the interpretation 'I visited many foreign countries, and now I am at home'), ends up as non-separative on the "surface", as required by the stative verb *somas* 'I was'.
13. In Sinti, the limitative particle *bis* is of course a loan from German. However, *bis* is also found in some Balkan dialects (e.g. Sofia Erli and Kosovo Bugurdži), where German origin can be safely excluded. We find it attractive to connect *bis* etymologically to the form *pos* of the same function (e.g. Kaspičan, Gadžikano, and Vălči Dol).
14. The origin of these forms is obscure. They are not explainable solely by phonological developments. The variant *kaj* may be a grammaticalisation of the place interrogative *kaj* 'where' (cf. Boretzky 2003: 38).
15. These markers are borrowed from the current contact languages, and may be considered to be parts of code-switches. In Šóka Rumungro, the use of Hungarian case markers is restricted to certain lexemes, and forms such as *Lenďelbe* 'in Poland' may be considered to be adverbs.
16. There is an example in Ajia Varvara with the Greek core preposition *se* 'in, into, at, to, on' used in the adessive localisation (*gelem* ***se*** *ekh bijav* 'I went to a wedding').
17. As other localisation-specific adpositions in the dialects, the loan forms a complex adposition with the general local (originally adessive) preposition *ke*: cf. *is ke*.
18. The ablative derivation *prek-al* was used in the perlative function in Bohemian Romani. However, the loan **preke* itself is unattested in this dialect.
19. The Macedonian preposition *karšija* is itself a loan from Turkish (*karşi*), also found in Greek (*karsi*).
20. There are also frequent loans of the local adverb 'away' (e.g. Slovene Romani *proč* from Slovene; Slovak Romani *het* from dialectal Slovak; Roman *bejg* and Austrian Sinti *veg* from German).

Chapter 18

1. Kibrik's (2003) term for our category of orientation is *direction*, based on Kilby's (1981) *directionality*.

2. The form *ko-te* in Gilan Arli is an interrogative rather than a deictic.
3. The pro-words in *-*ia* in some Central dialects (cf. the interrogative *k-ija* and the deictics, e.g. *ok-ija*) and the interrogatives *kev* (*t'ew*) in Slovene and Ukrainian Romani are of unknown origin, and may be old. In Varna Kalajdži, there is a mysterious non-separative form *kas-tar*, which is homonymous to the ablative of the person interrogative: 'where, whither' appears to be conceptualised as 'from whom'.
4. Šóka Rumungro *onď-ānde* 'inside STA' has been modelled on Hungarian *oda-benn* '(there) inside STA': Hungarian *oda*, like Romani *onďa*, is a directive deictic.

Chapter 19

1. From where we borrow the term *indefiniteness*.
2. There is also a simple de-interrogative variant *so-godi* 'whatever'.
3. This division of functions is not paralleled in person indefinites in Varna Kalajdži, where *k*(*h*)*onik* 'somebody, anybody, nobody' is a specific-to-negative indefinite.
4. As the prefix is not limited to dialects that show Rumanian influence elsewhere in their grammar or lexicon, it has probably diffused within Romani.
5. The extent of the shift depends on the ontological value in Italian Sinti and Abruzzian Romani (see Chapter 20).
6. The origin of some specific indefiniteness markers (e.g. *di-/dži-* in Slovene Romani, or *tirmi-* in Ursari) remains obscure and is likely to be due to grammaticalisation.
7. The Ukrainian prefix *de-* is also borrowed as such into Ukrainian Romani (see Section 19.5).
8. The same form in Nógrád Rumungro probably arose though a different process (viz. through extension of the oblique stem to the nominative), as the corresponding interrogative is *ko* 'who' rather than **kon*.
9. This can be assumed on the basis of the Iranian Romani form *hešta* 'nothing', which is best explained as a result of replacement of the South Slavic prefix *ni-* by the functionally equivalent Azeri/Persian prefix *hič-*: cf. **ni-šta* > **hič-šta* > *hešta*.
10. Itself a loan from Persian.
11. Loans of other semantic types of universal determiners are also common: e.g. *celo* (*calo*, *cilo*, *čilo*) from Slavic, *intrego* from Rumanian, or *krejt* from Albanian, all meaning 'whole'.
12. Apart from these, Xoraxane has borrowed the negative-polarity indefinite *išta* 'anything' from Serbian/Croatian, and Crimean Romani has borrowed the irrealis-negative polarity *šton'ebut'* 'something, anything' from Russian.
13. Cf. the derivation *har-tipda* 'everything'.
14. There is also the perlative local indefinite *valamerre* 'through somewhere' from Hungarian in some Rumungro varieties (see Chapter 18).
15. Perhaps the lack of the negative place indefinite in dialects in contact with Greek is due to its wide functional range in the source language (viz. it encodes negative polarity as well as negation).

16. Ultimately a loan from German.
17. Person indefinites do not present a clear picture, perhaps because they are partly homonymous to determiners.

Chapter 20

1. Diessel (1999) uses the term *demonstratives* to include demonstrative pronouns and determiners as well as demonstrative adverbs (of place and manner) and identifiers. Our term *deictics* is wider in that it covers further deictic modifiers (e.g. quality 'such as this', cardinal quantity 'this much', size 'this big') and further deictic adverbs (e.g. multiplicative quantity 'this many times'). We reserve the term *demonstratives* for demonstrative pronouns and determiners.
2. Only in a few dialects is the dative form of the thing interrogative distinct from the cause/goal interrogative (e.g. Bohemian Romani *soske* 'what.DAT' vs. *hoske* 'why').
3. We have little data on determiner-base manner and quantity indefinites.
4. Borrowed generic nouns are especially common with determiners borrowed from the same language (e.g. Muzikanta *vsjaku vreme* 'always; all the time' from Bulgarian, or Varna Gadžikano *er gün* 'always; every day' from Turkish). However, they are not restricted to such constructions (e.g. Yerli *bek vreme* 'sometimes; some time', or Muzikanta *kek vakəci* 'sometimes; some time').
5. The form *kan-ek* possibly developed from **kaj-ni-jekh* (see also Chapter 19); alternatively the prefix *kan-* is borrowed from Greek (see Section 19.5). The form *b-ek* might be a result of contamination of the forms *k-ek* or *d-ek* by the initial consonant of the Turkish-derived determiner *baazi* 'some'.
6. Only a minority of dialects possess a distinct class of relative pro-words, which are derived from interrogatives by means of a borrowed relative (and subordinative) marker: the Bulgarian suffix *-to* in many dialects of Bulgaria (e.g. Sofia Erli, Yerli, Varna Bugurdži, Muzikanta, Nange, Vlax Kalajdži, Rešitare; e.g. *kon-to* 'who REL'), or the Hungarian prefix *a-* in some Rumungro varieties (e.g. *a-savo* 'which REL'). Even in these dialects, the relative marker is not always obligatory, and so mere interrogatives may function as relativisers.
7. In Finnish Romani, most Central dialects, and Kosovo Bugurdži, the place interrogative appears to be restricted to relative clauses that modify a noun with a local meaning (such as 'place' or 'house'), and so it is better analysed as a connector in an embedded interrogative clause, rather than a relativiser proper.
8. We only analyse the most important types of adverbial subordination.
9. In some Balkan (e.g. Sofia Erli, Crimean Romani, Nange) and Vlax (e.g. Taikon Kalderaš, Rešitare, Kalburdžu) dialects, the place interrogative is also part of an anterior-durative subordinator (cf. *dži kaj*, *bis kaj*, *pos kaj* 'until'). This is less clear for the Vlax dialects, where the interrogative *kaj* coincides in form with an adessive preposition 'at, to', which may be a part of an anterior–durative case marker in some dialects (e.g. Rumungro *dži ko linaj* 'until the summer').

10. In Kaspičan, the causal subordinator consists of a cause/goal interrogative and a focus particle (*soske-da*), and in Nange, it consists of a cause/goal interrogative and a subordinative marker borrowed from Bulgarian (*soske-to*).
11. In most dialects, the connector is a particle: the standard of comparison is assigned the case relation of the comparee (e.g. subject *sar me* 'as I', direct object *sar man* 'as me'). In the dialects of Bulgaria, Macedonia, and Greece, the connector is a preposition that governs a particular case of the comparee nominal (e.g. *sar mande* 'as I/me').
12. In some of these dialects, there are equative connectors that derive from the general pool of quantity interrogatives, but are distinct from the actual interrogative (e.g. Nange *sibor* is only an equative connector, while *kizom* is an equative connector as well as the interrogative).
13. This hierarchy appears to be compatible with more precise (but also more arbitrary) rankings that would weight the position of individual ontological values in individual construction-dependent asymmetries. For example, if, for each construction, we assign 8 points to the most prominent value, 7 points to the second most prominent value etc. and zero points to values that do not show any extended distribution in that construction, we arrive at the following hierarchy: thing (31) > place (29) > manner (21) > quantity (19) > time (16) > cause/goal (8) > determiner (7) > person (5).
14. In Rumungro, these quantity interrogatives are specialised for size: e.g. Šóka Rumungro *kibedor* 'how big' (alongside the analytic *savo baro* 'how big') vs. *kiti* 'how much, how many'.
15. The counts are, by necessity, partly arbitrary (in that it is not always obvious what should count as an independent instance of borrowing) and partly inaccurate (since our data are not strictly comparable, i.e. there are gaps in data for some dialects).

Chapter 21

1. See Chapter 13 for a definition of this verb class.
2. Table 21.2 abstracts from dialect specific innovations in the shape of the forms. We disregard dialect types with a mixed encoding of the present in lexical verbs (see Table 13.3 in Chapter 13 for details).
3. It should be mentioned in this context that, in the majority of Romani dialects, pronouns show obligatory synthetic case marking in contexts where nouns do not (cf. Matras 1997:73).
4. Similar generalisations of various non-nominative pronominal forms took also place in English Romani and in a number of Para-Romani varieties (i.e. varieties that have retained only Romani-derived lexicon, which is used within the grammatical and discourse framework of another language; see Matras 2002, Ch. 10).
5. In Sepečides, non-emphatic pronominal genitives neutralise number of the possessor (e.g. *m-* 'my, our' vs. emphatic *mindr-* 'my' and *amar-* 'our').
6. Those Romani dialects which lack the regularised pronominal genitives altogether use the irregular genitives even in this construction (e.g. Italian Kalderaš *bi muřo*).

Chapter 22

1. Genitive associative forms are exceptional. They result from a morphological haplology of the expected forms (e.g. *lakatoš-i-n-ger-* < **lakatoš-i-n-ger-e-n-ger-*), and thus they are of equal complexity as the corresponding non-associative forms (cf. *lakatoš-e-n-ger-*).

Chapter 23

1. It should be noted that some borrowed grammatical markers do not participate in chronological compartmentalisation, i.e. not all borrowed markers are xenoclitic. First, syntactically free borrowed markers are never restricted to constructions with borrowed lexemes (although this is logically possible). Second, numerous bound borrowed markers apply to all grammatically relevant lexemes, without any regard to their origin (e.g. the attenuative suffix *-ast-* from Croatian in Rumungro).
2. Slavic adjectives are mostly adapted through inflections alone even in those dialects that possess overt adaptation markers for adjectives (e.g. Rumungro *žut-o* 'yellow' < Croatian *žut*). German adjectives are mostly unadapted (e.g. Sinti *hart* 'hard').
3. With an exception of a single dialect (see below).
4. The distinction between oikoclitic and xenoclitic masculines in *-o* is fully retained in Piedmontese Sinti.
5. There have been various person and number extensions of the suffix *-i* in some dialects (see Chapter 7 for details). Some Vlax dialects (e.g. Serbian Kalderaš, Rakarengo, Gurbet, and the South Vlax dialects of Bulgaria) do not show any remnants of this xenoclitic marker, and so they share the complete oikoclitic extension with the majority of other Romani dialects.
6. There is blurring of the compartmentalisation in Roman, whereas borrowed verbs may form oikoclitic participles (e.g. *pisin-do* alongside *pis-imo* < *pisin-* 'write') and indigenous verbs may form xenoclitic participles (e.g. *dikl-imo* alongside *dik-lo* < *dik-* 'see').
7. This stems from the diachronic logic of the phenomenon of chronological compartmentalisation, although it need not be a part of its formal definition.

Chapters 24–27

No notes

Appendix

1. The closely related Lajenge dialect, which was spoken in eastern Estonia until the extermination of all of its speakers, has not been sufficiently documented and is not included in our sample.

References

Andersen, Henning
1989 Markedness – the first 150 years. In *Markedness in Synchrony and Diachrony*, Olga M. Tomić (ed.), 11–46. Berlin: Mouton de Gruyter.
2001 Markedness and the theory of linguistic change. In *Actualization: Linguistic Change in Progress*, Henning Andersen (ed.), 21–58. Amsterdam: John Benjamins.

Andrews, Edna
1990 *Markedness Theory: The Union of Asymmetry and Semiosis in Language*. Durham, NC and London: Duke University Press.

Auer, Peter
1984 *Bilingual Conversation*. Amsterdam: John Benjamins.

Baghbidi, Hassan Rezai
2003 The Zargari language: An endangered European Romani in Iran. *Romani Studies*, Series 5, 13, 123–148.

Bailey, Charles-James N.
1973 *Variation and Linguistic Theory*. Arlington, Va.: Center for Applied Linguistics.

Bakker, Peter
1997 Athematic morphology in Romani: The borrowing of a borrowing pattern. In Matras, Bakker, and Kyuchukov (eds.), 1–21.
1999 The Northern branch of Romani: Mixed and non-mixed varieties. In Halwachs and Menz (eds.), 172–209.
2001 Typology of Romani numerals. *Sprachtypologie und Universalienforschung* 54: 91–107.

Bakker, Peter, and Yaron Matras
1997 Introduction. In Matras, Bakker, and Kyuchukov (eds.), vii–xxx.
2003 *A bibliography of modern Romani linguistics*. Amsterdam: John Benjamins.

Barannikov, Alexej P.
1934 *The Ukrainian and South Russian Gypsy Dialects*. Leningrad: Academy of Sciences of the USSR.

Bardovi-Harlig, Kathleen
1987 Markedness and salience in second-language acquisition. *Language Learning* 37(3): 385–407.

Battistella, Edwin L.
1996 *The Logic of Markedness*. New York and Oxford: Oxford University Press.

Bell, Alan
1978 Language samples. In *Universals of Human Language*, Vol. 1, Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik (eds.), 123–156. Stanford: Stanford University Press.

Berretta, Monica
1995 Morphological markedness in L2 acquisition. In *Iconicity in Language*, Rafaele Simone (ed.), 197–233. Amsterdam: Benjamins.

Bever, Thomas G., and D. Terrence Langedoen
1972 The interaction of speech perception and grammatical structure in the evolution of language. In *Linguistic Change and Generative Theory*, Robert Paul Stockwell and Ronald K. S. Macaulay (eds.). Bloomington: Indiana University Press.

Bickerton, Derek
1981 *Roots of Language*. Ann Arbor: Karoma.

Boretzky, Norbert
1986 Zur Sprache der Gurbet von Priština (Jugoslawien). In *Giessener Hefte für Tsiganologie* 3(1–4): 195–217, R. Gronemeyer, M. Muenzel, and G. A. Rakelmann (eds.).
1993a *Bugurdži. Deskriptiver und historischer Abriss eines Romani-Dialekts*. (Balkanologische Veröffentlichungen 21.) Wiesbaden: Harrasowitz.
1993b Conditional sentences in Romani. *Sprachtypologie und Universalienforschung* 46, 83–99.
1994 *Romani: Grammatik des Kalderaš-Dialekts mit Texten und Glossar*. Berlin: Harrassowitz Verlag.
1995 Die Entwicklung der Copula im Romani. *Grazer Linguistische Studien* 43: 1–50.
1996a Arli. Materialien zu einem südbalkanischen Romani-Dialekt. *Grazer Linguistische Studien* 46: 1–30.
1996b The "new" infinitive in Romani. *Journal of the Gypsy Lore Society*, Series 5, 6, 1–51.
1997 *Der Dialekt von Prilep*. Manuscript.
1998 Erli. Eine Bestandsaufnahme nach den Texte von Gilliat-Smith. *Studii Romani* 5–6: 122–160.
1999a *Die Verwandschaftbeziehungen zwischen den Südbalkanischen Romani-Dialekten. Mit einem Kartenanhang*. Frankfurt am Main: Peter Lang.
1999b Die Gliederung der Zentralen Dialekte und die Beziehungen zwischen Südlichen Zentralen Dialekten (Romungro) und Südbalkanischen Romani-Dialekten. In Halwachs and Menz (eds.), 210–276.
2000 South Balkan II as a Romani dialect branch: Bugurdži, Drindari, and Kalajdži. *Romani Studies*, Series 5, 10: 105–183.
2003 *Die Vlach-Dialekte des Romani. Strukturen – Sprachgeschichte – Verwandtschaftsverhältnisse – Dialektkarten*. Wiesbaden: Harrassowitz.

Boretzky, Norbert, and Birgit Igla
1994 *Wörterbuch Romani–Deutsch–Englisch für den südosteuropäischen Raum: mit einer Grammatik der Dialektvarianten*. Wiesbaden: Harrassowitz.

Borin, Lars
2000 A corpus of written Finnish Romani texts. In *LREC 2000. Workshop Proceedings. Developing Language Resources for Minority Languages: Reusability and Strategic Priorities*, Donncha O Croinin (ed.), 75−82. Athens: ELRA.

Bourgeois, Henri
1911 Esquisse d'une grammaire du romani finlandais. *Atti della Reale Academia delle Scienze di Torino* 46: 541−554.

Brøndal, Viggo
1940 Compensation et variation, deux principes de linguistique générale. *Scientia* 67: 101−109.

Bubeník, Vít, and Milena Hübschmannová
1998 Deriving inchoatives and mediopassives in Slovak and Hungarian Romani. *Grazer Linguistische Studien* 50: 29−44.

Campbell, Lyle
1980 Explaining universals and their exceptions. In *Papers from the Fourth International Conference on Historical Linguistics*, E. C. Traugott et al. (eds.), 17−26. Amsterdam: Benjamins.

Cech, Petra, and Mozes F. Heinschink
1998 *Basisgrammatik*. (Arbeitsbericht 1 des Projekts "Kodifizierung der Romanes-Variante der Österreichischen Lovara"). Wien: Verein Romano Centro.
1999 *Sepečides-Romani: Grammatik, Texte und Glossar eines türkischen Romani-Dialekts*. (Balkanologische Veröffentlichungen, 34.) Wiesbaden: Harrassowitz.
2001 A dialect with seven names. *Romani studies*, Series 5, 11: 137−184.

Chomsky, Noam, and Morris Halle
1968 *The Sound Pattern of English*. New York: Harper and Row.

Chvany, Catherine V.
1985 Backgrounded perfective and plot line imperfectives: toward a theory of grounding in text. In *The scope of Slavic Aspect*, Michael S. Flier and Alan Timberlake (eds.), 247−273. Columbus, Ohio: Slavica.

Comrie, Bernard
1981 *Language Universals and Linguistic Typology: Syntax and Morphology*. Oxford: Blackwell.
1986 Markedness, grammar, people, and the world. In *Markedness*, Fred R. Eckman, Edith A. Moravcsik, and Jessica R. Wirth (eds.), 85−106. New York and London: Plenum.
1993 Language universals and linguistic typology: Data-bases and explanations. *Sprachtypologie und Universalienforschung* 46: 3−14.

Corbett, Greville G.
1991 *Gender*. Cambridge: Cambridge University Press.
2000 *Number*. Cambridge: Cambridge University Press.
Crevels, Mily, and Peter Bakker
2000 External possession in Romani. In Elšík and Matras (eds.), 151–185.
Cristofaro, Sonia
2003 *Subordination*. Oxford: Oxford University Press.
Croft, William
1990 *Typology and Universals*. Cambridge: Cambridge University Press.
1995 Modern syntactic typology. In *Approaches to Language Typology: Past and Present*, Masayoshi Shibatani and Theodora Bynon (eds.), 85–143. Oxford: Oxford University Press.
2000 *Explaining Language Change: An Evolutionary Approach*. Harlow: Longman.
2003 *Typology and Universals*. 2nd ed. Cambridge: Cambridge University Press.
Dalton-Puffer, Christiane
1996 *The French Influence on Middle English Morphology*. Berlin and New York: Mouton de Gruyter.
DeGraff, Michel
1999 Creolisation, language change, and language acquisition: An epilogue. In *Language Creation and Language Change: Creolisation, Diachrony and Development*, Michel DeGraff (ed.). Cambridge, Mass.: MIT Press.
Diessel, Holger
1999 *Demonstratives: Form, Function, and Grammaticalisation*. (Typological Studies in Language 42.) Amsterdam: Benjamins.
Dixon, R. M. W.
1995 Complement clauses and complementation strategies. In *Grammar and Meaning: Essays in Honour of Sir John Lyons*, F. R. Palmer (ed.), 175–220. Cambridge: Cambridge University Press.
Djonedi, Fereydun
1996 Romano-Glossar: gesammelt von Schir-ali Tehranizade. *Grazer Linguistische Studien* 46: 31–59.
Dressler, Wolfgang, Willi Mayerthaler, Oswald Panagl, and Wolfgang U. Wurzel
1987 *Leitmotifs in Natural Morphology*. Amsterdam: Benjamins.
Dryer, Matthew S.
1989 Large linguistic areas and language sampling. *Studies in Language* 13: 257–292.
Eckman, Fred R.
1977 Markedness and contrastive analysis hypothesis. *Language Learning* 27(2): 315–330.

Elšík, Viktor
1997 Towards a morphology-based typology of Romani. In Matras, Bakker, and Kyuchukov (eds.), 23–59.
2000a Dialect variation in Romani personal pronouns. In Elšík and Matras (eds.), 65–94.
2000b Romani nominal paradigms: Their structure, diversity, and development. In Elšík and Matras (eds.), 9–30.
2000c Inherited indefinites in Romani. Paper presented at the Fifth International Conference on Romani Linguistics, Sofia, 14–17 September 2000.
2001a Word-form borrowing in indefinites: Romani evidence. *Sprachtypologie und Universalienforschung* 54: 126–147.
2001b Review of Halwachs 1998. *Romani Studies*, Series 5, 11: 53–66.
In prep. *A Grammar of Rumungro*.
Elšík, Viktor, and Yaron Matras (eds.)
2000 *Grammatical Relations in Romani: The Noun Phrase*. (Current Issues in Linguistic Theory 211.) Amsterdam: Benjamins.
Finck, Franz N.
1903 *Lehrbuch der Dialekts der deutschen Zigeuner*. Marburg: N. G. Elwert.
Frajzyngier, Zygmunt
1991 The de dicto domain in language. In *Approaches to Grammaticalisation*, Vol. 1, Elizabeth Closs Traugott and Bernd Heine (eds.), 219–251. Amsterdam: Benjamins.
Frajzyngier, Zygmunt, and Robert Jasperson
1991 *That* clauses and other complements. *Lingua* 83: 133–153.
Franzese, Sergio
1985 *Il dialetto dei Sinti Piemontesi. Note grammaticali. Glossario*. Torino: centro studi Zingari di Torino.
1986 *Il dialetto dei Rom Xoraxane. Note grammaticali. Glossario*. Torino: Centro Studi Zingari di Torino.
Fraser, Angus
1992 *The Gypsies*. Oxford: Blackwell.
Friedman, Victor
1985 Balkan Romani modality and other Balkan languages. *Folia Slavica* 7: 381–389.
1999 The Romani language in the Republic of Macedonia: Status, usage, and sociolinguistic perspectives. *Acta Linguistica Academiae Scientiarum Hungaricae* 46: 317–339.
Gilliat-Smith, B. J.
1915 A report on the Gypsy tribes of North East Bulgaria. *Journal of the Gypsy Lore Society*, New series, 9, 1–55, 65–109.
Givón, Talmy
1979 *On Understanding Grammar*. New York: Academic Press.

Givón, Talmy
1984 *Syntax. A Functional-Typological Introduction*. Vol. 1. Amsterdam: John Benjamins.
1990 *Syntax. A Functional-Typological Introduction*. Vol. 2. Amsterdam: Benjamins.
Givón, Talmy (ed.)
1983 *Topic Continuity in Discourse*. Amsterdam: John Benjamins.
Gjerde, Lars (with Knut Kristiansen)
1994 *"The Orange of Love" and Other Stories. The Rom–Gypsy Language in Norway*. Oslo: Scandinavian University Press.
Gjerdman, Olof, and Erik Lundberg
1963 *The Language of the Swedish Coppersmith Gipsy Johan Dimitri Taikon: Grammar, Texts, Vocabulary and English Word Index*. Uppsala: Lundequist.
Greenberg, Joseph H.
1966 *Language Universals, With Special Reference to Feature Hierarchies*. The Hague: Mouton.
1978 How does a language acquire gender markers? In *Universals of Human Language*, Vol. 3: *Word structure*, Joseph H. Greenberg, Charles A. Ferguson and Edith A. Moravcsik (eds.), 47–82. Stanford: Stanford University Press.
Gumperz, John
1982 *Discourse Strategies*. Cambridge: Cambridge University Press.
Haiman, John
1985 Iconic and economic motivation. *Language* 59(4): 781–819.
Halwachs, Dieter W.
1996 Verschriftlichung des Roman. *Grazer Linguistische Studien* 45: 1–32.
1998 *Amaro vakeripe Roman hi — Unsere Sprache ist Roman. Texte, Glossar und Grammatik der burgenländischen Romani-variante*. Klagenfurt: Drava.
Halwachs, Dieter W., and Florian Menz (eds.) 1999. *Die Sprache der Roma. Perspektiven der Romani-Forschung in Österreich im interdisziplinären und internazionalen Kontext*. Klagenfurt: Drava.
Hancock, Ian F.
1998 *The Indian Origin and Westward Migration of The Romani People*. Manchaca: International Romani Union.
2000 The emergence of Romani as a koïné outside of India. In *Scholarship and the Gypsy Struggle: Commitment in Romani Studies. A Collection of Papers and Poems to Celbrate Donald Kenrick's Seventieth Year*, Thomas Acton (ed.), 1–13. Hatfield: University of Hertfordshire Press.
Haspelmath, Martin
1993 More on the typology of inchoative/causative verb alternations. In *Causatives and Transitivity*, Bernard Comrie and Maria Polinsky (eds.), 87–120. Amsterdam: John Benjamins.

1997a *From Space to Time: Temporal Adverbials in the World's Languages*. (Lincom Studies in Theoretical Linguistics, 3.) Munich and Newcastle: Lincom Europa.
1997b *Indefinite Pronouns*. Oxford: Clarendon Press.
2002 *Understanding Morphology*. London: Arnold.

Heath, Jeffrey
1978 *Linguistic Diffusion in Arnhem Land*. (Australian Aboriginal Studies: Research and Regional Studies 13). Canberra: Australian Institute of Aboriginal Studies.

Hengeveld, Kees
1998 Adverbial clauses in the languages in Europe. In *Adverbial Clauses in the Languages in Europe*, Johan van der Auwera (ed.), 335–419. Berlin: Mouton de Gruyter.

Herzog, Marvin I., Uriel Weinreich, and Vera Baviskar (eds.) 1992. *The Language and Culture Atlas of Ashkenazic Jewry*. Vol. 1: *Historical and theoretical foundations*. Tübingen: Niemayer.

Himmelmann, Nikolaus
1997 *Deiktikon, Artikel, Nominalphrase: Zur Emergenz syntaktischer Struktur*. (Linguistische Arbeiten 362.) Tübingen: Niemeyer.

Hjelmslev, Louis
1935 *La catégorie des cas. Étude de grammaire générale*. (Acta Jutlandica 7.) Aarhus: Universitetsforlaget.

Holzinger, Daniel
1993 *Das Romanes: Grammatik und Diskursanalyse der Sprache der Sinte*. (Insbrucker Beiträge zur Kulturwissenschaft 85.) Innsbruck: Verlag des Instituts für Sprachwissenschaft der Universität Innsbruck.

Hübschmannová, Milena
1995 Trial and error in written Romani on the pages of Romani periodicals. In *Romani in Contact: The History, Structure and Sociology of a Language*, Yaron Matras (ed.), 189–205. Amsterdam: John Benjamins.

Hübschmannová, Milena, Hana Šebková, and Anna Žigová
1991 *Kapesní slovník romsko-český a česko-romský*. Prague: Státní pedagogické nakladatelství.

Hutterer, Miklós, and György Mészáros
1967 *A lovāri cigány dialektus leíró nyelvtana. Hangtan, szóképszés, alaktan, szótár*. Budapest: Magyar nyelvtudományi társaság.

Igla, Birgit
1996 *Das Romani von Ajia Varvara: deskriptive und historisch-vergleichende Darstellung eines Zigeunerdialekts*. Wiesbaden: Harrassowitz.

Jakobson, Roman
1932 Zur Struktur des russischen Verbums. In *Charisteria Gvilelmo Mathesio qvinqvagenario a discipulis et Circuli Lingvistici sodalibus oblata*, 79–84. Prague: Pražský lingvistický kroužek.

Jakobson, Roman
1936 Beitrag zur allgemeinen Kasuslehre: Gesamtbedeutungen der russischen Kasus. *Travaux du cercle linguistique de Prague* 6: 240–288.
1939 Signe zéro. In *Mélanges de linguistique offerts à Charles Bally*, 143–152. Geneva: Librairie de l'Université.
1941 *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist and Wiksells.
1957 *Shifters, Verbal Categories, and the Russian Verb*. Cambridge, Mass.: Russian Language Project, Harvard University.
1958 Typological studies and their contribution to historical comparative linguistics. In *Proceedings of the 8th International Congress of Linguists, 5–9 August, Oslo 1957*, 17–25, 33–35. Oslo: Oslo University Press.
Jakobson, Roman, and Linda Waugh
1979 *The Sound Shape of Language*. Bloomington: Indiana University Press.
Keesing, Roger
1988 *Melanesian Pidgin and the Oceanic Substrate*. Stanford: Stanford University Press.
Kemmer, Suzanne
1993 *The Middle Voice*. Amsterdam: John Benjamins.
Kenrick, Donald S.
1967 The Romani dialect of a musician from Razgrad. *Balkansko ezikoznanie* 11(2): 71–78.
Kibrik, Aleksandr E.
1991 Organising principles for nominal paradigms in Daghestanian languages: Comparative and typological observations. In *Paradigms: the Economy of Inflection*, Frans Plank (ed.), 255–274. (Empirical Approaches to Language Typology 9.) Berlin: Mouton de Gruyter.
2003 Nominal inflection galore: Daghestanian, with side glances at Europe and the world. In *Noun Phrase Structure in the Languages of Europe*, Frans Plank (ed.), 37–112. Berlin: Mouton de Gruyter.
Kilby, David
1981 On case markers. *Lingua* 54: 101–133.
König, Ekkehard, and Haspelmath, Martin
1997 Les constructions à possesseur externe dans les langues de l'Europe. In *Actance et valence dans les langues de l'Europe*, J. Feuillet (ed.), 525–606. Berlin: Mouton de Gruyter.
Koptjevskaja-Tamm, Maria
2000 Romani genitives in cross-linguistic perspective. In *Grammatical Relations in Romani: The Noun Phrase*, Viktor Elšík and Yaron Matras (eds.), 123–149. Amsterdam: Benjamins.
Kortmann, Bernd
1999 Typology and dialectology. In *Proceedings of the 16th International Congree of Linguists, Paris 1997*, B. Caron (ed.). CD-Rom. Amsterdam: Elsevier Science.

Kortmann, Bernd (ed.)
2004 *Dialectology Meets Typology*: *Dialect Grammar From a Cross-linguistic Perspective.* Berlin: Mouton de Gruyter.

Lípa, Jiří
1963 *Příručka cikánštiny*. Prague: Státní pedagogické nakladatelství.

Lüdtke, Helmut
1980 Sprachwandel als universales Phänomen. In *Kommunikationstheoretische Grundlagen des Sprachwandels*, Helmut Lüdtke (ed.), 182–252. Berlin: Mouton de Gruyter.

Lyons, John
1977 *Semantics*. Cambridge: Cambridge University Press.

McManus, Chris
2003 *Left Hand Right Hand. The Origins of Asymmetry in Brains, Bodies, Atoms and Cultures.* London: Weidenfeld and Nicolson.

Maddieson, Ian
1984 *Patterns of Sounds*. Cambridge: Cambridge University Press.

Mānušs, Leksa, Jānis Neilands, and Kārlis Rudevičs
1997 *Čigānu–latviešu–angļu etimoloģiskā vārdnīca un latviešu–čigānu vārdnīca*. Rīgā: Zvaigzne ABC.

Masica, Colin
1991 *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.

Matras, Yaron
1994a *Untersuchungen zu Grammatik und Diskurs des Romanes: Dialekt der Kelderaša/Lovara*. Wiesbaden: Harrassowitz.
1994b Structural Balkanisms in Romani. In *Sprachlicher Standard und Substandard in Südosteuropa und Osteuropa. Beiträge zum Symposium vom 12.–16. Oktober 1992 in Berlin*, Norbert Reiter, Uwe Hinrichs and Jiřina van Leeuwen-Turnovcová (eds.), 195–210. Wiesbaden: Harrassowitz.
1995 Verb evidentials and their discourse function in Vlach Romani narratives. In *Romani in Contact. The History, Structure and Sociology of a Language*, Yaron Matras (ed.), 95–123. Amsterdam: John Benjamins.
1997 The typology of case relations and case layer distribution in Romani. In Matras, Bakker, and Kyuchukov (eds.), 61–93.
1998a Deixis and deictic oppositions in discourse: Evidence from Romani. *Journal of Pragmatics* 29: 393–428.
1998b Utterance modifiers and universals of grammatical borrowing. *Linguistics* 36: 281–331.
1999a The speech of Polska Roma: some highlighted features and their implications for Romani dialectology. *Journal of the Gypsy Lore Society*, Series 5, 9, 1–28.
1999b The state of present-day Domari in Jerusalem. *Mediterranean Language Review* 11: 1–58.

Matras, Yaron
1999c S/h alternation in Romani: An historical and functional interpretation. *Grazer Linguistische Studien* 51: 99–129.
1999d Subject clitics in Sinti. *Acta Linguistica Academiae Scientiarum Hungaricae* 46(3–4): 147–169.
1999e Writing Romani: The pragmatics of codification in a stateless language. *Applied Linguistics* 20(4): 481–502.
2001 Tense, aspect, and modality categories in Romani. *Sprachtypologie und Universalienforschung* 53(4): 162–180.
2002 *Romani: A Linguistic Introduction*. Cambridge: Cambridge University Press.
2004 Romacilikanes: The Romani dialect of Parakalamos. *Romani Studies*, Series 5, 14: 59–109.
Matras, Yaron, Peter Bakker, and Hristo Kyuchukov (eds.) 1997. *The Typology and Dialectology of Romani*. (Current Issues in Linguistic Theory 156.) Amsterdam: John Benjamins.
Mayerthaler, Willi
1981 *Morphologische Natürlichkeit*. Wiesbaden: Athenäum.
1987 System-independent morphological naturalness. In Dressler et al. (eds.), 25–58.
Meier, Tinka
1999 Phonological markedness revisited: Evidence from Caucasian languages. In *Studies in Caucasian Linguistics*, Helma van den Berg (ed.), 181–196. Leiden: CNWS.
Mészáros, György
1976 The Cerhāri Gipsy dialect. *Acta Orientalia Academiae Scientiarum Hungaricae* 30: 351–367.
1980 *A Magyarországi szinto cigányok (történetük és nyelvük)*. Budapest: Magyar nyelvtudományi társaság.
Miklosich, Franz
1872–80 *Über die Mundarten und Wanderungen der Zigeuner Europas*. Vol. 3. Vienna: Karl Gerold's Sohn.
Mufwene, Salikoko
1990 Transfer and the substrate hypothesis in creolistics. *Studies in Second Language Acquisition* 12: 1–23.
1991 Pidgins, creoles, typology, and markedness. In *Development and Structures of Creole Languages: Essays in Honor of Derek Bickerton*, Francis Byrne and Thom Huebner (eds.), 123–143. Amsterdam: Benjamins.
Mühlhäusler, Peter
1980 Structural expansion and the process of creolisation. In *Theoretical Orientations in Creole Studies*, Albert Valdman and Arnold Highfield (eds.), 19–55. New York: Academic Press.

Müller, Friedrich
1869 Beiträge zur Kenntniss der Rom-Sprache. *Sitzungsberichte der philosophisch-historischen Classe der Kaiserlichen Akademie der Wissenschaften zu Wien* 61(1): 149–206.
Muysken, Pieter
1981 Creole tense–mood–aspect systems: the unmarked case? In *Generative Studies in Creole Languages*, Pieter Muysken (ed.), 181–199. Dordrecht: Foris.
Myers-Scotton, Carol
1993 *Dueling Languages: Grammatical Structure in Code-switching*. Oxford: Clarendon Press.
Nichols, Johanna
1992 *Linguistic Diversity in Space and Time*. Chicago: Chicago University Press.
Paspati, Alexandre G.
1973 [1870] *Etudes sur les Tchinghianés ou Bohémiens de l'empire ottoman*. Osnabrück: Biblio Verlag.
Payne, Doris L., and Immanuel Barshi (eds.)
1999 *External Possession.* (Typological Studies in Language 39.) Amsterdam: John Benjamins.
Perkins, Revere D.
1989 Statistical techniques for determining language sample size. *Studies in Language* 13: 293–315.
Plank, Frans
1995 (Re-)Introducing Suffixaufnahme. In *Double Case: Agreement by Suffixaufnahme*, Frans Plank (ed.), 3–110. New York: Oxford University Press.
Pobożniak, Tadeusz
1964 *Grammar of the Lovari dialect*. Kraków: Państwowe wydawnictwo naukowe.
Puchmayer, Anton Jaroslaw
1821 *Románi čib, das ist: Grammatik und Wörterbuch der Zigeuner Sprache, nebst einigen Fabeln in derselben. Dazu als Anhang die Hantýrka oder die čechische Diebessprache*. Prague: Fürst-erzbischöflichen Buchdruckerey.
Rijkhoff, Jan, Dik Bakker, Kees Hengeveld, and Peter Kahrel
1993 A method of language sampling. *Studies in Language* 17: 169–203.
Ross, Malcolm
1996 Contact-induced change and the comparative method: cases from Papua New Guinea. In *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*, Mark Durie and Malcolm Ross (eds.), 180–217. Oxford: Oxford University Press.

Ruhlen, Merritt
1987 *A Guide to the World's Languages*. Vol. 1: *Classification*. Stanford: Stanford University Press.
Sampson, John
1926 *The Dialect of Gypsies of Wales, Being an Older Form of British Romani Preserved in the Speech of the Clan of Abram Wood*. Oxford: Clarendon Press.
Sasse, Hans-Jürgen
1995 Theticity and VS-order: a case study. In *Verb–Subject Order and Theticity in European Languages*, Hans-Jürgen Sasse and Yaron Matras (eds.), 3–31. Berlin: Akademie. (Special issue of *Sprachtypologie und Universalienforschung* 48 (1–2)).
van Schooneveld, Cornelis
1978 *Semantic Transmutations. Prolegomena to a Calculus of Meaning*. Bloomington: Physsardt.
Shapiro, Michael
1972 Explorations into markedness. *Language* 48(2): 343–364.
Siegel, Jeff
1999 Transfer constraints and substrate influence in Melanesian Pidgin. *Journal of Sociolinguistics* 2: 347–373.
Silva-Corvalán, Carmen
1994 *Language Contact and Change: Spanish in Los Angeles*. Oxford: Clarendon.
Skalička, Vladimír
1979 *Typologische Studien*. Braunschweig and Wiesbaden: Viewegh.
Smart, Bath C., and Henry Thomas Crofton
1875 *The Dialect of the English Gypsies*. London: Asher and Co.
Song, Jae Jung
2001 *Linguistic Typology: Morphology and Syntax*. Harlow and London: Pearson Education.
Soravia, Giulio
1977 *Dialetti degli Zingari italiani*. (Profei dei dialetti italiani 22.) Pisa: Consiglio Nazionale delle Ricerche.
Soravia, Giulio, and Camillo Fochi
1995 *Vocabolario sinottico delle lingue zingare parlate in Italia*. Roma: Centro Studi Zingari.
von Sowa, Rudolf
1887 *Die Mundart der slovakischen Zigeuner*. Göttingen: Vandenhoeck und Ruprechts Verlag.
Sperber, Dan, and Dierdre Wilson
1986 *Relevance: Communication and Cognition*. Oxford: Blackwell.
Spinelli, Santino
n.d. Unpublished manuscript on Abruzzian Romani (1990s).

Stassen, Leon
1985 *Comparison and Universal Grammar*. Oxford: Blackwell.
Stein, Dieter
1989 Markedness and linguistic change. In *Markedness in Synchrony and Diachrony*, Olga M. Tomić (ed.), 67–85. Berlin: Mouton de Gruyter.
Talmy, Leonard
1983 How language structures space. In *Spatial Orientation: Theory, Research and Application*, Herbert L. Pick and Linda P. Acredolo (eds.), 225–282. New York: Plenum Press.
Thomason, Sarah Grey, and Terrence Kaufman
1988 *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
Tiersma, Peter M.
1982 Local and general markedness. *Language* 58(4): 832–849.
Tomlin, Russell S.
1986 *Basic Word Order: Functional Principles*. London: Croom Helm.
Trubetzkoy, Nikolai S.
1939 *Grundzüge der Phonologie*. (Travaux de circle linguistique de Prague 7.)
Trudgill, Peter
1989 Contact and isolation in linguistic change. In *Language Change: Contributions to the Study of Its Causes*, Leiv Egil Bervik and Ernst Håkon Jahr (eds.), 227–237. Berlin: Mouton de Gryuter.
Valet, Joseph
1991 Grammar of Manush as it is spoken in the Auvergne. In *In the Margins of Romani: Gypsy Languages in Contact*, Peter Bakker and Marcel Cortiade (eds.), 106–131. Amsterdam: Institute for General Linguistics.
Valtonen, Pertti
1972 *Suomen mustalaiskielen etymologinen sanakirja*. Helsinki: Suomalaisen kirjallisuuden seura.
Vekerdi, József
1971 The Gurvari Gypsy dialect in Hungary. *Acta Orientalia Academiae Scientiarum Hungaricae* 24: 381–389.
1984 The Vend Gypsy dialect in Hungary. *Acta Linguistica Academiae Scientiarum Hungaricae* 34: 65–86.
Vincent, Nigel
1978 Is sound change teleological? In *Recent Developments in Historical Phonology*, Jacek Fisiak (ed.), 409–430. The Hague: Mouton.
Waugh, Linda
1982 Marked and unmarked: A choice between unequals in semiotic structure. *Semiotica* 38: 299–318.

Weinreich, Uriel
1953 *Languages in Contact: Findings and Problems*. New York: Linguistic Circle of New York.
Wentzel, Tatjana
1980 *Die Zigeunersprache (Nordrussischer Dialekt)*. Leipzig: Enzyklopädie.
Wierzbicka, Anna
1988 The semantics of English complementation in a cross-linguistic perspective. In *The Semantics of Grammar*, Anna Wierzbicka (ed.), 23–168. Amsterdam: Benjamins.
Windfuhr, Gernot
1970 European Gypsy in Iran: A first report. *Anthropological Linguistics* 12: 271–292.
Winford, Donald
2003 *An Introduction to Contact Linguistics*. Oxford: Blackwell.
Winter, Werner
1989 Markedness and naturalness. In *Markedness in Synchrony and Diachrony*, Olga M. Tomić (ed.), 103–109. Berlin: Mouton de Gruyter.
Wurzel, Wolfgang U.
1987 System-dependent morphological naturalness in inflection. In Dressler et al. (eds.), 59–97.
1989 Inflectional class markedness. In *Markedness in Synchrony and Diachrony*, Olga M. Tomić (ed.), 227–248. Berlin: Mouton de Gruyter.
2000 Inflectional system and markedness. In *Analogy, Levelling, Markedness: Principles of Change in Phonology and Morphology*, Aditi Lahiri (ed.), 193–214. Berlin: Mouton de Gruyter.

Index of authors

Andersen, Henning, 8, 9
Andrews, Edna, 8
Auer, Peter, 42

Baghbidi, Hassan Rezai, 415
Bailey, Charles-James N., 23
Bakker, Peter, 55–57, 68, 162, 165, 166, 324, 370, 429
Barannikov, Alexej P., 130, 412, 418
Bardovi-Harlig, Kathleen, 26
Barshi, Immanuel, 219
Battistella, Edwin L., 7, 9–12, 16
Baviskar, Vera, 54
Bell, Alan, 48, 50
Berretta, Monica, 26
Bever, Thomas G., 23
Bickerton, Derek, 24
Boretzky, Norbert, 54–57, 94, 127, 135, 201, 203, 255, 300, 301, 414–417, 429, 432, 435
Borin, Lars, 411
Bourgeois, Henri, 411
Brøndal, Viggo, 9
Bubeník, Vít, 432

Campbell, Lyle, 23, 24
Cech, Petra, 55, 127, 193, 244, 246, 300, 414–416
Chomsky, Noam, 9
Chvany, Catherine V., 8
Comrie, Bernard, 13, 17, 48–50
Corbett, Greville G., 138, 322
Crevels, Mily, 55
Cristofaro, Sonia, 181
Croft, William, 4, 10–12, 15–20, 23, 25, 28, 35, 39, 43, 48, 49
Crofton, Henry Thomas, 410

Dalton-Puffer, Christiane, 26
DeGraff, Michel, 24
Diessel, Holger, 54, 437
Dixon, R. M. W., 60
Djonedi, Fereydun, 415
Dressler, Wolfgang, 13, 15, 23, 24
Dryer, Matthew S., 48–50

Eckman, Fred R., 11, 24
Elšík, Viktor, 48, 55, 56, 68, 72, 76, 77, 86, 94, 99, 104, 125, 140, 142, 200, 229, 234, 282, 284, 311, 317, 318, 320, 329, 413, 416, 425, 429, 432

Finck, Franz N., 411
Fochi, Camillo, 416
Frajzyngier, Zygmunt, 60
Franzese, Sergio, 411, 417
Fraser, Angus, 68
Friedman, Victor, 203, 425

Gilliat-Smith, B. J., 56
Givón, Talmy, 4, 10, 12, 17, 18, 24, 45, 60, 181, 425
Gjerde, Lars, 416
Gjerdman, Olof, 244, 416
Greenberg, Joseph H., 10–12, 15–17, 19, 43, 44, 54
Gumperz, John, 42

Haiman, John, 16
Halle, Morris, 9
Halwachs, Dieter W., 55, 246, 413, 425
Hancock, Ian F., 425
Haspelmath, Martin, 16, 18, 73, 259, 281, 288, 290, 432
Heath, Jeffrey, 26
Heinschink, Mozes F., 55, 127, 193, 244, 246, 300, 411, 414–416

Hengeveld, Kees, 181
Herzog, Marvin I., 54
Himmelmann, Nikolaus, 54
Hjelmslev, Louis, 8
Holzinger, Daniel, 55, 182, 299, 411
Hübschmannová, Milena, 413, 425, 432
Hutterer, Miklós, 416

Igla, Birgit, 55, 94, 135, 244, 415, 417, 432

Jakobson, Roman, 7–11, 25
Jasperson, Robert, 60

Kaufman, Terrence, 23–26
Keesing, Roger, 24
Kemmer, Suzanne, 432
Kenrick, Donald S., 416
Kibrik, Aleksandr E., 432–435
Kilby, David, 435
König, Ekkehard, 73
Koptjevskaja-Tamm, Maria, 55, 73
Kortmann, Bernd, 50, 53
Kyuchukov, Hristo, 68

Lípa, Jiří, 413
Lüdtke, Helmut, 23
Lyons, John, 18

Maddieson, Ian, 49
Mānušs, Leksa, 412, 429
Masica, Colin, 71
Matras, Yaron, 33, 42, 52–57, 68, 74, 79, 82, 84, 99, 107, 118, 140, 174, 181–188, 199, 203, 208, 215, 218, 370, 394, 396, 410, 412, 416, 425, 438
Mayerthaler, Willi, 18, 22
McManus, Chris, 1
Meier, Tinka, 16
Mészáros, György, 257, 411, 416
Miklosich, Franz, 55, 56, 69
Mufwene, Salikoko, 26
Mühlhäusler, Peter, 24, 25
Müller, Friedrich, 413
Muysken, Pieter, 24
Myers-Scotton, Carol, 3

Neilands, Jānis, 412, 429
Nichols, Johanna, 24

Paspati, Alexandre G., 56, 414, 415
Payne, Doris L., 219
Perkins, Revere D., 49
Plank, Frans, 71, 221
Pobożniak, Tadeusz, 416
Puchmayer, Anton Jaroslaw, 127, 413

Rijkhoff, Jan, 50
Ross, Malcolm, 42
Rudevičs, Kārlis, 412, 429
Ruhlen, Merritt, 49

Sampson, John, 235, 261, 410
Sasse, Hans-Jürgen, 186, 187

Šebková, Hana, 413

Shapiro, Michael, 8, 9
Siegel, Jeff, 26
Silva-Corvalán, Carmen, 42
Skalička, Vladimír, 14
Smart, Bath C., 410
Song, Jae Jung, 48
Soravia, Giulio, 411, 414, 416
Sperber, Dan, 46
Spinelli, Santino, 414
Stassen, Leon, 49
Stein, Dieter, 23, 24, 27

Talmy, Leonard, 434
Thomason, Sarah Grey, 23–26
Tiersma, Peter M., 12, 13, 26
Tomlin, Russell S., 49, 50
Trubetzkoy, Nikolai S., 7, 10
Trudgill, Peter, 25

Valet, Joseph, 269, 411, 430
Valtonen, Pertti, 411
van Schooneveld, Cornelis, 8
Vekerdi, József, 413
Vincent, Nigel, 23
von Sowa, Rudolf, 413

Waugh, Linda, 8
Weinreich, Uriel, 26, 42, 54
Wentzel, Tatjana, 412
Wierzbicka, Anna, 60
Wilson, Dierdre, 46
Windfuhr, Gernot, 415
Winford, Donald, 25, 26
Winter, Werner, 18
Wurzel, Wolfgang U., 20, 21

Žigová, Anna, 413

Index of Romani dialects

Abruzzian Romani, 95, 104–106, 113, 123, 124, 142, 167, 168, 171, 192, 193, 200, 212, 234, 235, 284, 289, 292, 293, 309, 310, 318, 320, 401, 414, 431, 436
Ajia Varvara, 88, 93, 94, 110, 113, 118, 133, 135, 146, 148, 153, 165, 179, 185, 191, 196, 197, 199–201, 214, 231, 236, 237, 243, 244, 250, 254, 265, 266, 268, 274, 275, 278, 284, 286, 290–293, 304–307, 309, 317, 402, 417, 431, 435
Arli:
  Florina, 63, 93, 94, 110, 113, 120, 121, 133, 144, 146, 148, 153, 159, 165, 172, 179, 191, 197, 207, 209, 210, 214, 231, 236, 237, 243, 248, 250, 260–263, 265, 270, 277, 283, 284, 293, 296, 298, 299, 304–307, 309, 313, 314, 317, 414, 431, 434, 435
  Gilan, 94, 97, 103, 106, 110, 113, 118, 120, 127, 133, 151, 191, 193, 209, 214, 236, 237, 265, 277, 278, 289, 309, 310, 318, 330, 414, 427, 431, 434, 436
  Karditsa, 105, 107, 152, 153, 171, 205, 208, 210, 214, 234, 236, 237, 248, 291, 292, 301, 314, 414, 430, 434
  Kumanovo, 151–153, 156, 165, 172, 200, 210, 214, 225, 236, 237, 252, 254, 260, 262, 265, 292, 293, 311, 313, 414, 434
  Prilep, 87, 93, 94, 97, 110, 113, 118, 120, 125, 132, 133, 139, 152, 159, 161, 165, 178, 179, 185, 191, 194, 196–198, 206, 209, 214, 230, 231, 234, 236, 243, 244, 248, 254, 260–262, 267–270, 289, 292, 293, 305–307, 309–311, 313, 316, 317, 414, 427, 431, 434
  Prizren, 107, 110, 111, 142, 146, 153, 165, 168, 171, 190, 191, 193, 198, 201, 237, 253, 284, 287, 289, 292, 304–307, 317, 414, 427, 431
  Skopje, 103, 106, 110, 118, 151, 156, 191, 193, 230, 314, 414, 427, 431

Bohemian Romani, 127, 150, 169, 178, 180, 209, 266, 285, 303, 306, 331, 434, 435, 437

Calabrian Romani, 318, 414
Catalonian Romani, 284, 285
Cerhari, 99, 106, 147, 151, 171, 201, 416, 417
Čergare *see* Rešitare
Crimean Romani, 93, 94, 100, 110, 113, 120, 125, 133, 135, 136, 144, 147, 150, 152, 153, 165, 168, 171, 172, 179, 191, 205, 214, 215, 248, 254, 276, 284, 289, 291, 292, 301, 305–307, 312, 331, 414, 436, 437

Dasikano, 94, 106, 107, 110–113, 125, 132, 133, 146, 150, 168, 184, 191, 200, 207, 237, 268, 274, 289, 291, 292, 309, 310, 316, 317, 417, 431

Early Romani, 5, 6, 28, 34, 37, 51, 52, 54, 62, 66, 68, 70, 72–77, 80, 81, 83, 85–92, 94–99, 101, 103, 105, 107, 108, 111, 113, 115, 116, 121, 122, 125–127, 132, 134, 138–140, 142–144, 146, 157, 159, 160, 163–167, 171, 176, 183, 184, 186, 190, 192, 194, 196–202, 211, 2`15, 216, 218,

221–223, 225–229, 233, 234, 241, 242, 253, 259, 260, 266, 273, 275, 279, 281, 282, 284, 302, 305, 306, 309, 312, 315, 316, 318, 324, 325, 330, 332, 391, 396, 397, 426, 429, 431, 433
English Romani, 410, 411, 438
Epiros Romani, 35, 82, 128, 150, 212, 414, 427
Estonian Romani, 111, 164, 165, 168, 172, 247, 255, 256, 258, 283, 307, 310, 412

Finnish Romani, 94, 95, 106, 107, 109, 110, 113, 115, 116, 118, 122, 128–130, 142, 146, 147, 154, 161, 163, 165, 168, 177, 179, 181–183, 186, 190, 193, 194, 199, 200, 205, 214, 215, 236, 238, 255, 262, 276, 277, 279, 284, 285, 291–293, 298, 302, 304, 306, 314, 315, 318, 320, 330, 411, 426, 427, 431, 433, 437

Gurbet:
- Bačka, 171, 200
- Kumanovo, 151–153, 156, 165, 172, 200, 210, 214, 225, 236, 237, 252, 254, 260, 262, 265, 292, 293, 311, 313, 414, 434
- Priština, 103, 113, 118, 146, 150, 164, 191, 199, 214, 225, 253, 270, 284, 293, 306, 309, 310, 417
- Srem, 171, 200
- Vojvodina, 171

Gurvari, 289, 290, 292, 318, 412

Kalburdžu, Sindel , 87, 94, 101, 103, 110, 136, 139, 144, 158, 165, 175, 180, 183, 186, 191, 192, 199–201, 207, 214, 229, 230, 237, 243, 245, 253, 257, 261, 265, 279, 304, 306, 307, 309, 394, 418, 429, 431, 437
Kalderaš
- Bunkuleš, 104, 110, 113, 164, 172, 191, 228, 270, 276, 284, 285, 428, 429
- Italian, 5, 75, 94, 171, 235–237, 277, 293, 310, 317, 414, 417, 436, 438
- Markuleš, 110, 123, 416
- Serbian, 63, 87, 94, 103, 104, 110, 113, 122, 123, 127, 133, 141, 142, 144, 146, 160, 163–165, 171, 172, 190, 191, 194–196, 198, 207, 214, 228, 231, 237, 244, 246, 254, 255, 261–263, 270, 276, 277, 279, 283–286, 289, 291–293, 296, 298, 300, 302, 304, 309–311, 313, 315, 317, 319, 320, 331, 332, 416, 417, 428, 429, 435, 437–439
- Taikon, 110, 113, 122, 127, 133, 141, 144, 163, 165, 171, 190, 195, 196, 214, 228, 231, 244, 246, 254, 262, 263, 270, 277, 279, 283, 286, 291, 293, 298, 309, 315, 319, 331, 332, 416, 417, 435, 437

Kaale, 411
Kāle, 410
Kaspičan, 101, 103, 110, 135, 136, 139, 147, 161, 163, 170, 177–179, 183, 191, 199, 201, 207, 235–237, 254, 265–267, 276, 277, 279, 291, 293, 296, 301, 304, 305, 310, 321, 410, 415, 435, 438
Kelderaš, see Lovari
Kosovo Bugurdži, 63, 93, 94, 110, 113, 117, 127, 144, 147, 151, 159, 164, 165, 168, 180, 199, 200, 202, 209, 214, 230, 231, 236, 237, 243, 244, 246, 251, 254, 260, 269, 276, 285, 286, 289, 291–293, 298, 304, 306, 307, 309–311, 321, 396, 415, 434, 435, 437

Lajenge 439
Latvian Romani, 35, 88, 90, 110, 133, 150, 159, 167, 191, 194, 251, 275,

Latvian Romani (*cont.*)
276, 279, 286, 291, 310, 318, 330, 332, 412
Lithuanian Romani, 65, 98, 128, 135, 147, 150, 160, 165, 168, 180, 198, 205, 206, 209, 210, 229, 231, 234, 237, 247, 257, 260–262, 265, 267, 269, 270, 279, 293, 312, 317, 319, 412, 428, 433
Lom, 200, 201, 276
Lovari, 63, 87, 94, 95, 97, 105, 106, 110, 113, 114, 117, 118, 122, 123, 127, 129, 133, 141, 142, 144, 147, 151, 158, 159, 165, 171, 172, 174–177, 179, 184, 185, 187, 191, 194–196, 198, 199, 207, 213, 214, 216, 228, 231, 244, 246, 251, 254, 261, 262, 264, 268, 277, 278, 291, 293, 300, 302, 306, 307, 309, 310, 313, 314, 316, 317, 330–332, 412, 416, 417, 426–428, 433, 434
Austrian, 53, 59, 87, 90, 94, 106, 113, 117, 118, 124, 127, 159, 161, 165, 168, 172, 190, 195, 196, 198, 199, 207, 213, 214, 231, 235, 236, 244, 246, 258, 261, 269, 277, 278, 283, 291, 300, 303–307, 310, 316, 317, 330, 426, 427, 430, 434, 435
Bougeští, 114, 115, 117, 196, 416, 417
Hungarian , 101, 292
Norwegian, 332, 417
Polish, 5, 35, 64, 91, 97, 104, 111, 113, 123, 129, 142, 147, 150, 160, 161, 165, 168, 172, 178, 180, 185, 191, 195, 204, 206, 207, 209, 214, 229, 231, 237, 245, 247, 255, 260, 265, 270, 276, 285, 291, 293, 298, 302, 306, 307, 310, 314, 317, 402, 403, 410, 412, 431, 433, 434

Montana Kalajdži, 151, 156, 165, 172, 254, 265, 276, 277, 289, 291–293, 410

Nange 93, 94, 100, 110, 133, 135, 139, 148, 165, 168, 183, 191, 199, 202, 207, 209, 214, 231, 237, 244–248, 254, 257–260, 263, 265, 269, 276, 277, 289, 292, 305, 307, 309, 314, 415, 437, 438

Pazardžik Malokonare, 160
Podolie Romani, 90, 191, 195
Polish Romani, 35, 64, 91, 104, 111, 129, 142, 147, 150, 160, 161, 165, 178, 180, 204, 206, 209, 229, 231, 245, 247, 255, 260, 265, 270, 276, 293, 298, 302, 306, 307, 314, 410, 412, 431, 433, 434
South Polish Romani, 91, 412, 433
Prekmurje Romani, 289
Proto-Romani, 52, 79, 84, 95, 142, 183, 212

Rakarengo, 163, 165, 200, 254, 263, 317, 416, 439
Razgrad Drindari, 94, 125, 293, 319
Rešitare, 93, 94, 106, 110, 113, 133, 148, 151, 160, 165, 168, 180, 184, 191, 200, 202, 206, 210, 214, 231, 235–237, 244, 254, 258, 260, 261, 263, 265, 269, 284, 302, 305–307, 309, 317, 321, 417, 437
Roman, 59, 63, 110, 123, 150, 157, 159, 165, 169, 172, 178, 180, 182, 185, 195, 198, 207, 209, 213, 214, 231, 237, 246, 251, 268, 269, 277, 306, 307, 316, 317, 329, 331, 413, 427, 429, 434, 435, 439
Romano, 142, 172, 414
Rumelian Romani, 63, 97, 110, 114, 118, 120, 132, 146, 153, 165, 172, 179, 195, 214, 304, 309, 320, 414, 426, 430
Rumungro:
Klenovec, 110, 128, 165, 177–179, 208, 209, 213, 215, 246, 261, 262,

266, 269, 270, 302, 306, 314, 317, 331, 413
Nógrád, 86, 101, 198, 215, 228, 234, 314, 436
Sípos, 139, 163, 165, 179, 206, 214, 219, 255, 260, 314, 330, 413, 429
Šóka, 87, 88, 110, 125–127, 129, 142, 151, 163, 165, 172, 179, 180, 189, 198, 199, 201, 212, 214, 216, 245, 250, 251, 258, 262, 264, 266, 267, 270–274, 277, 278, 280, 287, 290, 291, 295, 296, 312, 322, 330, 426, 432, 434–436, 438
Russian Romani, 63, 111, 146, 167–169, 171, 172, 190, 191, 231, 233, 261, 270, 277, 310, 316, 320, 331, 412, 429

Sepečides, 63, 87, 93, 94, 100, 106, 110, 113, 118, 125, 132, 133, 144, 146–148, 151, 153, 159, 160, 165, 168, 171, 172, 179, 191, 197, 201, 209, 213, 214, 234, 243, 244, 248, 260, 261, 263, 277, 286, 292, 293, 298, 304–306, 309, 311, 313, 327, 331, 399, 414, 433, 435, 438
Sinti:
Austrian, 53, 59, 87, 90, 94, 106, 113, 117, 118, 124, 127, 159, 161, 165, 168, 172, 190, 195, 196, 198, 199, 207, 213, 214, 231, 235, 236, 244, 246, 258, 261, 269, 277, 278, 283, 291, 300, 303–307, 310, 316, 317, 330, 426, 427, 430, 434, 435
Bohemian, 87, 110, 123, 124, 127, 142, 150, 154, 165, 169, 178, 180, 209, 266, 285, 303, 306, 314, 331, 412, 428, 434, 435, 437
German, 5, 18, 56, 63, 90, 119, 123, 124, 142, 149, 153, 159, 161, 165, 170, 172, 179, 185, 195, 213, 235, 236, 249, 251, 268–270, 277, 278, 286, 288, 289, 291–293, 306, 329, 330, 402, 403, 411–413, 428–430, 435, 437, 439
Hameln, 119, 153, 160, 165, 169, 230
Hungarian, 5, 59, 90, 91, 95, 101, 106, 113, 124, 129, 142, 150, 151, 154, 159–161, 164, 165, 167, 168, 170, 172, 179, 185, 186, 193, 195, 202, 214, 216, 217, 228, 234, 235, 251, 257, 264, 267–270, 277, 278, 285, 286, 288–293, 310, 322, 327, 329, 393, 402, 403, 412, 413, 417, 426, 430, 432–437
Lalere, 123
Lombardian, 106, 171, 317, 320
Manuš, 91, 92, 118, 119, 159, 165, 169, 172, 192, 194, 199, 231, 235, 236, 246, 268, 269, 291, 293, 411
Piedmontese, 106, 109, 123, 127, 146, 148, 150, 157, 159, 160, 165, 167, 169, 182, 190, 236, 277, 293, 306, 309, 310, 315, 317, 318, 329, 411, 427, 439
Slovak Romani:
Central Slovak Romani, 139, 140, 282, 289, 412
East Slovak Romani, 87, 92, 98, 103, 105, 110, 133, 140, 141, 159, 168, 171, 193, 198, 200, 276, 278, 286, 296, 327, 332, 396, 412
West Slovak Romani, 108, 110, 142, 159, 161, 169, 180, 209, 269, 270, 412, 429
Slovene Romani, 93, 94, 97, 99, 101, 103, 106, 109, 129, 133–135, 142, 144, 147, 150, 158, 160, 161, 164, 168, 170, 172, 179, 184, 186, 191–193, 197, 198, 201, 202, 204, 207, 209, 210, 213, 214, 229, 230, 236, 237, 244, 245, 254, 255, 260, 261, 267–270, 276–278, 285, 287, 289–293, 305–307, 310, 313, 317, 319, 320, 330, 401, 410, 414, 428, 429, 431, 434–436

Sofia Erli, 93, 94, 106, 110, 113, 114, 121, 127, 142, 144, 151, 156, 157, 164, 165, 168, 172, 175, 179, 186, 187, 191, 195, 199, 206, 214, 231, 235, 237, 238, 244, 245, 254, 259, 260, 265, 269, 276–278, 284, 286, 289, 292, 300, 302, 304–307, 309, 310, 313, 429, 435, 437
Šumen Drindari, 168, 172, 179, 181, 202, 207, 209, 301

Ukrainian Romani, 90, 94, 95, 103, 106, 124, 129, 130, 155, 286, 306, 309, 310, 330, 401, 418, 428, 431, 436

Varna Bugurdži, 93, 94, 110, 133, 139, 151, 163, 165, 169, 180, 209, 214, 231, 236, 237, 243, 245, 247, 248, 254, 259, 260, 265, 266, 269, 276, 277, 292, 293, 300, 301, 305–307, 309, 313, 314, 317, 414, 434, 437
Varna Gadžikano, 110, 265, 437
Varna Kalajdži, 93, 94, 105, 110, 113, 133, 135, 148, 157, 161, 165, 168, 172, 177, 179, 184, 191, 200, 207, 214, 230, 231, 243–245, 254, 257, 258, 260, 265, 269, 277–279, 283, 284, 301, 305–307, 309, 410, 417, 436
Vend, 142, 227, 413
Vidin Cocomanya, 150, 152, 153, 200
Vălči Dol, 109, 113, 118, 135, 136, 163, 165, 179, 191, 196, 199–201, 206, 209, 214, 236, 237, 253, 262, 263, 270, 284, 300, 302, 305–307, 417, 433, 435

Welsh Romani, 35, 37, 75, 87, 88, 93, 98, 100, 106, 110, 114, 117, 123, 127, 133, 142, 144, 146, 148, 157, 161, 167, 168, 171, 172, 185, 186, 191, 194, 196, 198–200, 209, 215, 216, 231, 235, 243, 246, 247, 260, 261, 265, 267, 274, 282, 284–286, 291, 298, 304, 305, 329, 396, 397, 410, 411, 426, 433

Xoraxane, 87, 97, 100, 103, 107, 110, 111, 113, 118, 123, 133, 142, 146, 150, 157, 160, 168, 191, 192, 197, 237, 265, 282, 289, 291, 292, 305–307, 309, 417, 431, 436

Yerli:
  Rakitovo, 163
  Velingrad, 41, 61, 62, 66, 163, 254, 259, 266, 270, 306, 307, 427

Zargari, 120, 139, 145, 155, 157, 169, 183, 309, 320, 410, 414, 429

Index of geographical names

Abruzzi, 414
Albania, 414, 417
Anatolia, 69
Ararat, 69
Armenia, 69
Austria, 13, 56, 123, 129, 146, 411, 413, 414, 417
Azerbaijan, 425

Balog (Slovakia), 170, 412
Balkans, 52, 55–58, 63, 68–70, 75, 84, 101, 113, 123, 133, 135, 150, 153, 160, 162, 171, 190–193, 203, 204, 207, 214, 230, 231, 235, 236, 243, 245, 252, 253, 260, 263, 266, 269, 278, 292, 298, 301, 305–307, 309, 313, 314, 317, 320, 327, 330, 397, 401–402, 410, 414, 433
Baltics, 123
Banat, 123, 194
Belarus, 412
Bohemia, 107
Bosnia, 417
Bratislava, 251
Britain, 410, 412
Bulgaria, 56, 58, 61, 144, 147, 151, 156, 168, 183, 186, 202, 204, 206, 237, 258, 259, 267, 269, 276, 291, 410, 414–417, 425, 437, 438, 439
Burgenland, 59, 150, 165, 169, 413
Bystrany (Slovakia), 430
Byzantium, 5, 69

Calabria, 414
China, 412
Croatia, 53, 57, 417
Czech Republic, 150, 288, 403, 411, 412, 413, 417, 425, 426, 429

England, 410
Estonia, 412, 425, 439

Finland, 411, 425
France, 123, 411

Georgia, 414
Germany, 13, 57, 123, 129, 411, 412, 414, 417
Greece, 69, 82, 414, 417, 425, 438
Helsinki, 109, 128, 129, 147, 154, 181, 185, 186, 206, 237, 238, 262, 284
Hungary, 56, 123, 129, 131, 411, 413, 425

Iran, 69, 414
Istria, 186, 414
Italy, 56, 100, 124, 410, 411, 414, 417
Izmir (Turkey), 151

Kiev, 130
Krompachy (Slovakia), 165, 171
Kuopio (Finland), 142, 183, 285, 292

Latvia, 412
Lithuania, 412, 425
Łódż (Poland), 65
Lučivná (Slovakia), 100, 128, 165, 178, 180, 230, 231, 237, 249, 258, 262, 263, 266, 299, 306, 412

Macedonia, 56, 60, 147, 151, 152, 156, 225, 233, 414, 415, 417, 425, 438
Moldavia, 416
Molise, 414
Montenegro, 417

Pabianice (Poland), 65

Podhalie (Poland), 412
Podhradie (Slovakia), 165, 170
Podunajské Biskupice (Slovakia), 131
Poland, 56, 65, 105, 125, 226, 237, 267, 269, 410, 412, 425, 435
Prague, 7, 10, 14, 19
Prekmurje (Slovenia), 413
Pribylina (Slovakia), 170, 412

Rumania, 289, 414, 415, 416, 417, 425
Russia, 123, 230, 235, 267, 412, 414, 418
Ruthenia (Ukraine), 412

Salgótarján (Hungary), 131
Serbia, 59, 414, 417, 425
Slavjansk (Ukraine), 130
Sliven (Bulgaria), 100, 278, 301, 415
Slovakia, 56, 105, 107, 125, 131, 154, 226, 251, 306, 330, 412, 413, 425, 428
Slovenia, 56, 413, 414
Spoitori (Rumania), 415
Štítník (Slovakia), 330
Švedlár (Slovakia), 170

Sweden, 243, 411

Transylvania, 53, 57, 58, 412, 416
Turkey, 147, 291, 414, 417

Ukraine, 56, 123, 206, 412, 414, 418

Vojvodina, 53, 57
Volos (Greece), 146, 148, 153

Wales, 410
Wallachia, 56, 57, 416, 417
Warsaw, 269

Zbojné (Slovakia), 170
Zborov (Slovakia), 165, 412
Zemplín (Slovakia), 98, 114, 115, 196

Index of subjects

ability, 157–161, 203–204, 207–209, 335, 338, 343, 354, 382, 389, 394, 398
  ability particle, 160, 428
accusative, 76, 92–93, 131, 175, 220–223, 226, 243, 247, 299, 301–303, 327, 335, 338, 427, 432 n. 2, 433 n. 7
  accusative clitic, 94
  accusative enclitic pronoun, 212
  vs. accusative nominative, 94
  reduced accusative form, 427 n. 7
  see also oblique
acquisition, 10, 22, 24
  child language, 2, 9, 11, 44
  order, 425 n. 2
  second language, 24, 26, 44
actualisation patterns, 27
adaptation markers, 58, 143, 199, 200, 211, 215, 217, 234, 326, 330, 396, 397, 439
adessive, 231, 239, 240–269, 434 n. 4, 435 n. 11, 16, 437 n. 9
  locative, 244
  non-separative adessive, 243
  separative adessive, 235, 243, 248, 253–255, 261–263, 265
adjectivals *see* adjective
adjective, 74, 86, 87, 101, 140–141, 143, 145, 149, 167, 172, 182, 211, 218, 229, 234, 270, 312, 315, 319, 325–326, 331, 332, 336, 338, 343, 390–396, 428, 439 n. 2
  vs. adverb, 149
  asymmetry, 336, 390
  consonantal, 87, 148, 163, 220, 227–228
  de-nominal, 318
  de-substantival, 326, 332
  Greek-derived, 74
  oikoclitic, 314, 327, 328
  privative, 156
  vocalic, 148
  xenoclitic, 95, 101, 329, 332, 433 n. 9
  see also borrowing
adverbials, 149, 181, 203, 235, 270, 306, 307, 353, 428, 437
  subordination, 84, 181, 203, 306, 307
  superlative marker, 428 n. 1 (ch. 9)
adverb, 68, 145, 156, 193, 211, 235, 248–252, 260, 266, 273–280, 331
  vs. adjective, 149
  vs. case markers, 220, 240, 435 n. 15
  core, 249, 251
  de-adjectival, 326, 332
  de-nominal, 318
  demonstrative/deictic, 437 n. 1
  derived from xenoclitic adjective, 332
  local, 83, 140, 239–241, 249, 250, 267, 271–274, 279, 434 n. 1, 5, 435 n. 11, 20
  phasal, 156, 157
affirmative, 156–160, 209, 368, 369, 372, 373, 376, 381, 384, 385, 390
  borrowing, 386
  modal, 389, 394
  verbs, 394
aktionsart, 188, 189, 193, 194, 334, 336, 337, 338, 339, 341, 342, 343, 368, 432
  modification, 189, 193, 202, 212, 379, 385
  neutral, 335, 338, 342, 343, 373
animacy, 71, 224, 247, 296, 301, 329, 380–381
  animacy continuum, 299

animacy (*cont.*)
animacy split, 300, 302
and oikoclitic–xenoclitic distinction, 329
polarised animacy scale, 299
anterior, 240, 242, 247, 251, 252, 258, 263, 265, 266, 267, 274, 437
anterior-durative, 259, 261–262, 437 n. 9
definite, 41, 68, 74–76, 84, 100, 150, 153, 173, 174, 182–184, 227, 236, 319, 335, 338, 343, 396
grammaticalisation, 146–148
indefinite, 174, 182–183, 184, 298, 399
origin, 53–54
relation to demonstrative, 54, 74, 183, 396
aspect, 55, 68, 82–83, 85, 102, 215, 334, 339, 348, 351, 376, 379, 380, 388, 389, 393, 396
inflection, 74, 140–141, 228, 319, 433 n. 9
markedness, 9
person, 57
voice, 34
associative, 322, 323, 439
associativity, 322, 334, 336, 337, 339, 341, 342, 343
non-associative, 322, 323, 335, 338, 439
asymmetry, 33, 220, 229, 308, 311, 379, 383, 388
(external) case, 224–225, 227, 319, 323, 354
(in)definiteness, 283, 286, 290, 311, 341, 354
adverbial–core case role, 232
cardinality, 166–167
categorial, 151
complexity asymmetry, 104, 156, 189, 193, 212, 221, 248–249, 283, 298, 334–336, 398–399, 428 n. 3
criteria, 29–32, 35–47, 279, 334–346, 354, 377, 406–409
degree, 145, 146–149, 153–154, 354
differentiation, 337–339, 399–400, 428 n. 1 (ch. 9)
diversity, 167, 169–170
erosion, 85, 156, 192, 250, 336–337, 346
extension, 339–340, 401
functionality in change, 406–409
gender, 140–141, 143, 336, 339
hierarchies, 347–351, 356–357, 362, 406
implicational, 237, 293
in borrowing, 160–161, 211, 290, 292, 343, 370, 385, 401–402
internal case, 232, 354
modality, 354
negators, 158
number, 88, 91–92, 100, 101, 226
orientation, 354
paradigmatic approach, 394–395
vs. participle finite forms, 89–90
perfective–imperfective, 393–394
person, 107, 116, 120–121, 130, 135–136, 195
pronouns, 86, 292, 303
remote–non-remote, 192
separative–non-separative, 254, 434 n. 7
stative–directive, 273
TAM marking, 120
tense, 194, 354
transitive–intransitive, 215, 216
types of substantival, 225
athematic classes *see* xenoclitic classes
auxiliarity, 312, 334, 336, 337, 339, 341, 342, 343, 391, 397
auxiliary, 57, 79, 160, 190, 193, 202, 335, 343, 368, 369, 395, 396, 397
non-auxiliary, 338, 368, 375, 397
axis localisation, 239, 241–242, 249, 251–252, 259, 264, 266–267, 270

borrowing:
ability modals, 161
adjectives, 143, 172, 234, 327, 329
adpositional case markers, 235
adpositions, 236, 237
adverb, 193, 267–270
affirmative, 386
affixes, 143
and asymmetry, 161
asymmetry, 211
case markers, 234, 235, 267–268
concord
conjunctions, 185
constraints, 26, 151, 154, 374
contrastive conjunction, 33, 34
degree markers, 151, 154, 428 n. 1 (ch. 9)
determiner, 196, 291, 310–311, 344, 372, 436 n. 11, 437 n. 4
from Bulgarian, 186, 234, 237
from Croatian, 320
from German, 179, 185, 236
from Greek, 54, 69, 71, 101, 171, 172, 179, 237, 320–321, 324, 430 n. 15
from Hungarian, 179, 185, 193
from Macedonian, 237
from Polish, 185, 237
from Rumanian, 144, 186, 234, 310
from Russian, 185, 237
from Serbian, 179, 282, 321
from Slavic, 144, 186
from Slovak, 237
from Slovene, 179, 401
from Turkish, 101, 185, 186, 321
grammatical, 42
indefinites, 293, 311
interrogative determiner, 310
interrogatives, 277, 309–311
localisation, 235, 371, 372,386
and markedness, 25–27, 41, 370–372, 409
motivations for, 389
negative auxiliary, 160
negative-polarity marker, 282
negator, 161, 200, 354
nominative markers, 234, 372
nouns, 26, 234
number markers, 101, 102, 344, 401–402
place indefinites, 292–293
plural markers, 101
polarity, 436 n.12
proneness/susceptibility to, 59, 62, 102, 146, 162, 173–174, 184, 185, 276, 321, 344, 381, 383, 386–387
pronouns, 277–278, 287–294, 310, 402
relative pronouns, 437 n. 6
replicative borrowing, 153, 154
typological distance, 26
valency-changing markers, 216
verbs, 320
YES-particle 429 n. 8 (ch. 10)

cardinality, 162, 166–167, 170, 334, 336, 337, 340–344, 349, 351, 367–369, 373–375, 379, 381, 385, 386, 388, 397, 404
ordinal, 369, 371, 373, 375
see also asymmetry
case
complexity asymmetries, 221, 222
differentiation asymmetries, 224–225
differentiation, 89, 93, 120, 121, 140, 224–229
double case, 71
erosion, 37, 52, 317–319, 336, 337–339
external, 218, 221, 224–226, 229, 233, 334, 354
homonymy, 92–94, 120
in adjectivals, 219
indigenous adpositions, 252
inflectional, 219, 247
internal, 218, 220–221, 224, 229, 232, 234, 334, 354, 360, 392

case (*cont.*)
irregularity, 121
retention, 52, 395
vocative, 218
see also borrowing
case I, 334, 389, 390, 392
case II, 334
case marking, 71, 92, 93, 175, 220, 221, 225, 234, 239, 243, 248, 250, 251, 299, 301, 438
adverbial, 232, 240
vs. adverbs, 240, 250
animacy, 71, 299
localisation, 243
non-separative, 260–261
reverse, 234
separative anterior, 274
separative, 262, 265
sociative, 229, 236
stative/directive, 278, 279
case roles, 39, 219–224, 232–238, 259, 334, 336, 337, 339, 341–343, 433 n. 3
adverbial + separative, 223
adverbial, 219–220, 222–223, 232–235, 238
associative, 223
vs. core adverbial, 219–220, 232
exceptive, 237, 238
extension, 229–230
separative, 233, 234, 250
substitutive, 237, 238
see also complexity; asymmetry
causal relations, 307
cause, 45, 219, 295–299, 302–308, 310, 337–344, 357–368, 378, 382, 400, 437, 438
change, external, 20, 44, 47, 23, 44, 45, 51, 52, 55, 58, 60, 62, 66, 85, 113, 347, 356, 372, 395, 406, 431, 432
complexity susceptibility to change, 51, 366–370
and diversity, 43
functionality of asymmetry, 406–409
internal, 23–25, 269
predictability, 11, 20
sound change, 197, 430–431 n. 3, 7
comparative, 29, 41, 51, 53, 55, 65, 68, 145–155, 165, 172, 266, 281, 307, 326, 335–343, 354, 358, 379, 429, 433, 434
complementiser, 39, 60–62, 126, 178, 179, 181, 206, 207, 306, 307, 398, 428
factual–non-factual split, 68, 84
non-factual, 61, 157, 160, 190
reversal, 222
conceptual motivation, 6, 30, 32–34, 44, 47, 347, 372, 377, 385, 388, 393
conditional, 22, 46, 55, 84, 115, 117, 185, 188, 189, 196, 203–205, 208–210, 281, 307, 315
conditionality, 203–205, 334, 336, 337, 339, 341–343, 354, 376, 381
connector, 164–166, 168, 178, 179, 203, 307, 429, 437, 438
constructional iconicity, 14, 18
degree of, 25
and markedness, 24
convergence, 42, 142, 184, 230, 231,282, 306, 396
adpositions, 236, 255, 433 n. 17
case roles, 232–234
comparative, 153
and Early Romani Greek, 84, 184, 305
future particle, 190
interrogative, 305
loss of gender distinction, 226
negative future, 156
structural, 84, 149, 153, 277, 286
with genderless languages, 318, 396
with Slavic languages, 231, 246, 286, 306, 319, 433 n. 17
with Ukrainian, 286
copula, 57, 105–107, 111, 126, 127, 156, 157, 159, 160, 188–190, 192–194,

198, 213, 234, 312–317, 320, 391, 395, 397, 429, 431, 433
core grammar, 10
core localisation, 239, 241, 242, 244, 245, 247, 249, 252, 255, 256, 259, 264, 265, 267, 268, 362, 369, 380, 381, 386, 434
Core Sinti, 106, 111, 127, 146, 147, 159, 190, 192, 216, 236, 268, 270, 274, 278, 285, 302, 329–330, 428 n. 1 (ch. 10), 433 n. 9

degree, 29, 41, 145–151, 153–155, 337, 340, 342, 354
  non-positive, 368, 386
  and number, 358, 365, 367, 379
deictics, 437 n. 1
demonstrative, 36, 54, 74–76, 82, 95, 98, 139–142, 159, 174, 183, 184, 227, 228, 287, 309, 316, 319, 390, 393, 426, 437
determiner, 78, 148, 284–285, 287, 297, 309, 337, 338, 340, 384, 398, 400
  deictic, 74
  distibutive, 291
  inflection, 303
  vs. quality identification, 296
  universal, 74
  see also article, pronoun
directive, 77, 83, 231, 240–242, 246, 248–252, 254, 261, 265, 272–280, 296, 338, 339, 341, 343, 379, 382, 391, 434, 436
discreteness, 32, 173–176, 186, 187, 204, 208, 334, 336, 337, 339, 341–343, 360, 364, 368, 370, 375, 380, 386, 404, 408
distributional potential, 19

equative constructions, 307
erosion, 36–37
evidentiality, 118, 194, 196, 208 334, 336–339, 341–343, 393, 394

factuality, 179, 334, 341–343
feminine factuality *see* gender
form–function correlation, 2, 12, 26, 41, 188
future, 37, 62–63, 91, 118, 119, 157, 188–194, 199, 202, 231, 232, 261, 263, 264, 315, 335–339, 341, 343, 354, 379, 380, 398, 408, 431 n. 8
  negative future, 156–157
  proclitic particle, 190

gender, 37, 40, 71–72, 82, 86, 95–96, 117, 121, 131, 138–144, 184, 224–225, 314, 319, 381, 390, 392–393, 395, 396, 407, 428 n. 2 (ch. 8), 433 n. 8
  default masculine, 360
  differentiation, 226
  discourse-prominence, 384
  and gender split in dative genitive, 225, 233
  neutralisation, 228, 319, 326, 426 n. 8
  retention, 314, 318
  vs. substantival adjectival, 303
  see also asymmetry
genitive, 71, 72, 76, 91, 105, 125, 130, 132, 133, 219, 220–226, 230–233, 243, 297, 302, 309, 314, 318, 320, 322–323, 354, 362, 364, 368, 395, 433 n. 8
  agreement, 71, 73
  attribute, 82, 84, 174
  cardinals, 164
  erosion, 37, 337
  marking split with dative, 225
  neutralisation, 228
  pronominal, 319
  pronouns, 76, 98, 105–106, 116, 125
  singular-to-plural extension, 98–99
goal/Goal, 77, 218, 295–297, 299, 302–305, 307, 308, 310, 335–340, 342–344, 382, 400, 437 n. 2, 438 n. 10

hierarchies, 12, 35, 85, 102, 108, 145, 149, 167, 259, 264, 267, 287, 338, 344–358, 360, 362, 364, 366, 368, 371, 375, 377, 379, 380, 384, 385, 387–395, 401, 402, 405–407
historical independence, 50
  preterite–pluperfect
homonymy, 426 n. 9
  case homonymy, 92, 94, 121, 219, 221
  comparative, 149
  Early Romani
  gender, 96, 226, 393
  grammatical, 8
  indefinite pronouns–article
  nominative–accusative, 92, 327
  nominative–oblique, 432 n. 1
  number, 96, 100, 113–118, 122, 129, 130, 140–141, 400, 426 n. 9, 427 n. 4, 5
  perfectivity, 113, 330
  person, 89–91, 194–196, 400, 427 n. 5, 8
  preterite–pluperfect, 192, 195, 426 n. 5
  subjunctive inflection, 118–119
  TAM, 91, 119–120, 330, 400
horizontal adpositions, 257, 263, 267

iconicity, 4, 8, 15, 17, 18, 377–383, 385, 408
imperative, 12, 85, 103, 125, 126, 158, 188–190, 194, 196–198, 200–202, 315, 335, 337, 338, 340, 350, 354, 382, 390, 431
imperfect, 26, 37, 90, 111, 119, 120, 124, 134, 141, 181, 188–190, 192–194, 202, 334–336, 339, 341, 343, 398, 430, 431
inability, 157–161, 335, 338, 343, 354, 382, 394, 428
  clausal complement, 430 n.2
  vs. conditional clauses, 208–209
  copula, 106–107, 198, 315, 316
  future, 190
  negators, 200–202
  past, 189
  present, 63, 158, 161, 203–204, 360–361, 369, 383, 430, 431
  and subject clitics, 198
  vs. subjunctive, 158, 204, 315, 337, 350, 354, 382, 431 n. 9
inessive, 239, 240, 242, 243–245, 248–262, 265, 267–268, 337, 434 n. 4, 435 n. 11
infinitive, 40, 61, 84, 125, 127, 129, 130, 179, 181, 202, 203, 204, 207, 428
inflectional class stability, 20, 21
irrealis, 77, 115, 117, 196, 203–205, 282, 284, 285, 290, 292, 338, 343, 345–346, 354, 381, 387, 436 n. 12

lexical specification, 21
lexicality, 140, 312–321, 352, 375, 385, 387, 389–391, 395, 397
  vs. auxiliarity non-auxiliarity, 312, 395–396
  hierarchy, 371, 387, 396
  lexicality hierarchy, 371
  lexicality, 1, 391, 395
  lexicality, 2, 334, 396
loan-shift, 42
loan-words *see* borrowing
localisation, 230–232, 235, 239–270, 271–275, 280, 337, 341, 352–353, 433 n. 3, 434 n. 1, 5, 7
  borrowing, 235, 371, 372, 386
  vs. case roles, 39, 243–249
  extension, 274
  groups of, 239, 241–242, 353, 360
  homonymy, 241
  human, 231
  sequentive, 231, 241
  superior, 230, 241

manner, 147, 295–297, 299, 302–303, 306–308, 311, 335–338, 340,

342–344, 357, 359, 361, 363–367, 373–375, 382, 400, 438
as a binary relation, 7–8
assimilation, 9
and borrowing, 368, 370, 372
consistency problem, 10
core grammar, 9
criteria, 11
extralinguistic correlates, 9, 17, 22, 41
implicational universals, 11
Jakobsonian markedness, 7–8, 17
vs. linguistic–extralinguistic, 13
vs. linguistic–situational, 13
local, 12–13, 26
markedness assimilation, 9
Markedness Differential Hypothesis, 11, 24
Markedness Hypothesis, 4, 355–365, 368, 370–371, 379, 404, 407–408
Markedness Model of Codeswitching, 3
markedness reversal, 9, 12
markedness shift, 23
vs. phonological semantic, 8
and pidgins creoles, 26
Praguean theory, 13
typological markedness, 10, 11, 12, 16, 19, 20
unmarkedness, 8
masculine, 39–40, 41, 71, 72, 95, 98–100, 131, 138–144, 148, 221, 224–229, 287, 318–319, 322, 327, 336, 339, 340, 360–362, 369, 381, 384, 390, 392–393, 396, 428
mental accessibility, 358, 407
metatypy, 42
defective, 125–126
modality, 55, 65, 82–83, 180–181, 186, 204, 208, 209, 335, 354, 382, 384, 389, 398–399
dependent, 203–204
vs. epistemic complementation, 60
modal conjunction, 203
obligative, 126
modifier, 71, 163, 173, 315, 362, 382, 395–396, 429
adjectival, 73–74, 84, 319
deictic, 437 n. 1
grammatical, 45
indeclinable, 303, 429 n. 1
modal, 203
mood, 91, 126, 158, 204, 201, 205, 204–205, 337, 343, 334, 350, 354, 360, 361, 369, 382

Natural Morphology, 13, 18
and language change, 20
Natural Phonology, 14
naturalness, 10, 12–15, 20–23, 25
extralinguistic foundations, 14
naturalness principles, 14–15
necessity, 17, 126, 127, 203, 207–209, 335, 338, 354, 382, 398, 429, 438
negation, 203, 185–186, 394–395, 389–390, 209, 281–287, 290, 316, 351, 381, 384, 386, 389, 390, 394, 395, 402, 428 n. 1 (ch. 10), 429 n. 6 (ch. 10)
subjunctive, 158, 431 n. 9
neutralisation, 7, 8, 12, 19, 396, 426 n. 8
gender, 95, 142, 228, 319, 322, 396
number, 112–116, 226, 228, 392–393
person, 124, 195
specific–plain, 184
TAM values, 158
nominative, 73–74, 143, 175, 218–223, 354, 428 n. 15
vs. accusative, 92–94, 299
vs. oblique, 52, 116, 228, 247, 358, 392–393, 400, 433 n. 9
see also borrowing, homonymy, internal case
non-perfective, 83, 89, 90, 91, 96, 97, 108, 118, 122, 124, 188, 189, 191–196, 198, 199, 200, 215, 316, 330, 335, 368, 375, 379, 380, 387, 393,

non-perfective (*cont.*)
394, 396, 397, 431 n. 3, n. 8
innovation, 134
subjunctive, 190
see also perfective
non-remote, 91, 93, 118–119, 188, 190, 192, 194–196, 354, 368, 372, 375, 381, 385
see also asymmetry
non-signalisation, 7, 17
noun, 71–72, 73, 218, 219, 220, 224, 247–248, 265–266, 322, 437 n. 7
abstract, 331
vs. adverb, 235
generic, 298, 305, 399
inflection, 227, 318, 328, 331, 433 n. 8
and modifier, 395–396
and pronoun, 320
see also borrowing, lexicality 2
number, 37, 39–40, 53, 71, 82, 104–105, 136, 140, 219, 319, 352, 365, 367–368, 379, 386–387, 390, 392, 407, 439 n. 5
vs. associativity, 322
differentiation, 114–117, 194–196, 224, 226, 337, 399, 433 n. 13
distinction, 112, 303, 392
hierarchies, 358
split, 51
see also asymmetry, homonymy, neutralisation

oblique, 72–74, 75–76, 93, 99, 100, 105, 120, 130, 131, 163, 175, 176, 218–229, 243, 304, 310, 316, 318, 322, 329, 330, 354, 360, 380, 400, 426, 429 n. 1, 433 n. 8, 436
animacy, 175, 301, 381
erosion, 37
and gender distinctions, 142–144, 227, 228, 319
gender neutralisation, 95
inflectional assimilation, 433 n. 10
number differentiation, 117
see also homonymy, internal case, nominative, renewal
oikoclitic classes, 72–75, 86, 87, 92, 95, 97, 134, 140, 196, 227–229, 314, 324–333, 335, 341–343, 433, 439
see also adjective, vowel systems
orientation, 70, 83, 248, 254, 257, 266, 268, 271–280, 303, 336, 341–343, 354, 368, 379, 383, 388, 391–392, 400–401, 434 n. 4, 435 n. 1
distinctions, 240, 251–252, 338
erosion, 265

paradigm structure conditions, 21
perfective, 89, 90, 92, 99, 119, 121, 135, 181, 188–190, 192, 194–200, 204, 211, 215, 216, 312–314, 316, 330–332, 339, 343, 379, 390, 393, 401, 402
classes, 34–35, 80, 91
concord, 81–82
formation, 34, 103
vs. nonperfective, 82, 394, 396–397, 430 n. 1
person–number marker, 96, 103–105, 113–115, 117, 121, 124, 133–134, 196, 426 n. 2
see also borrowing, homonymy
perlative, 83, 231, 240–246, 249, 258, 267, 270, 272, 273, 276–278, 280, 296, 338, 435, 436
person, 2–7, 29, 34, 37, 40, 77, 89, 90, 102, 108, 124, 131–135, 195–197, 298, 308, 310, 336, 338–341, 345, 348, 357, 359, 361, 363–367, 373–375, 378, 380, 384, 390, 427, 437
phonology, 7, 11, 12, 20, 35, 52, 55, 174
place, 177, 184, 268, 284–285, 287, 296, 297, 305, 308, 310–311, 316, 335, 338, 344, 357, 360–361, 369, 374, 400–401, 434 n. 5, 438 n. 13

deixis, 177, 184, 316
place interrogatives, 268, 305–306, 309–311, 435 n. 14, 437 n. 7, 9
pro-words, 303
see also borrowing
pluperfect, 83, 87–88, 98, 113–115, 117–118, 121–124, 181, 188–190, 192, 195–196, 315, 334, 339, 343, 426 n. 2, 3
see also homonomy, neutralisation
plural, *see* number
positive *see* degree
possessive, 38, 76, 77, 105, 174, 219, 234, 314, 338, 343, 395–396
possessive *see* genitive
posterior, 239, 240, 242, 247, 250–252, 257, 259, 262, 263, 265–267
potential, 7, 19, 20, 34, 38, 39, 45, 46, 48–50, 53, 92, 149, 173, 188, 189, 203–205, 315, 335, 338, 341, 343, 345, 346, 349, 354, 369, 381, 384–386, 388, 403, 408
inflectional potential, 16, 19, 38, 203, 393
preference structure, 9, 10
present, 52, 188–194, 315, 317, 334–337, 438 n. 2
present-future, 37, 62, 63, 118, 119, 190, 194, 202, 315, 336
present–subjunctive, 40, 63, 134, 135, 199
see also borrowing
preterite, 87, 91, 96, 113–115, 118, 121, 124, 134, 136, 160, 188–190, 192, 195, 196, 213, 214, 334, 339, 343, 426 n. 2
adverbial, 303
pro-words *see* pronouns
pronouns:
asymmetry criteria, 354–355
cause/goal indefinite, 310–311
cause/goal interrogative, 437 n, 1
combinations of lexial types, 298–299
de-indefinite, 283
free-choice indefinite, 77–78, 281–291, 296, 340–341, 344, 354, 368, 375, 380, 384, 402–404
immediacy hierarchies, 380, 383
indefinite place, 284, 285, 287, 292, 293, 305, 311, 436
indefinite quantity, 319
indefinites based on interrogative, 398–399
indefinite, 185, 200, 276, 278, 283–285, 287, 290–294, 296, 298, 303–305, 310, 311, 319, 380, 398, 399, 402, 436 n. 1
interrogative as subordinator, 305–308, 437 n. 6
vs. interrogative indefinite, 284, 286, 296, 399
interrogative indefinite, 74, 283, 297, 299
interrogatives as connectors, 307
interrogative, 37, 39–40, 74, 77, 99, 143, 271–272, 276–278, 283, 286, 295–296, 297, 299, 302, 316, 317, 403, 438 n. 14
interrogative, 37, 77, 143, 226, 276, 297, 303, 398–399
innovation, 277
manner indefinite, 293
manner interrogative, 37, 302, 306–307, 310
negative indefinite, 39, 78, 200, 278, 283–287, 289, 293, 294, 304, 311, 436 n. 9, 15
negative-polarity indefinite, 436 n. 12
origin of free-choice pronouns, 283
perlative indefinite, 278, 436 n. 14
personal, 76–77, 86, 93, 102, 105, 124–125, 183, 226, 247, 318; case differentiation 89, 120; gender differentiation 89, 121, 340, 396; and demonstratives 176, 182–183; loss of case 318; person differentiation

pronouns (*cont.*)
91, 93, 112, 336; and reflexive pronouns 130-131, 428 n. 15
reflexive, 76, 99, 100, 125, 130–133, 143, 318, 428 n. 15
relative, 39, 84, 305, 306, 437
specific indefinite, 281, 283, 287, 290, 291, 293, 294, 311, 436 n. 6
specific-to-negative, 77–78, 281, 283–285, 436
substantival, 99, 143, 303
universal, 77–78, 282–287, 293, 294, 304, 361, 362, 369, 380, 402, 436 n. 13
see also asymmetry, borrowing
prototype, 12–13, 386, 390
proximate, 9, 75, 173, 182, 239, 240, 242, 247, 249–251, 256, 257, 259, 264–268, 270, 434

quantity, 74, 293, 295–297, 299, 303, 306–311, 319, 335, 338, 340, 343, 344, 358, 368, 369, 371–375, 377, 379, 381–385, 387, 392, 393, 400, 405, 409, 437, 438
question, 281–282
see also pronouns/interrogatives

realis, 204, 205, 209, 210, 335, 341, 343, 345, 346, 354, 369, 372, 376, 381, 385, 387
reduplication, 43, 176, 177, 182, 278, 283, 286, 304
remoteness marker, 83, 91, 87, 108, 111–112, 116, 176, 181, 192, 193, 199, 364, 398, 408, 427 n. 6, 430–431 n. 3
see also asymmetry
renewal, 30, 43, 273, 275–276, 346, 368, 408
case roles, 220, 234
clitics, 144
copula, 316
degree, 154
demonstratives, 58
gender, 143, 144
internal case, 232, 393
interrogative, 57
modals, 160
number, 387, 390
oblique form of interrogative, 57
polarity, 159, 160
pronouns, 395
separative, 273, 276, 278
third person, 34
representability, 49, 50

sample, 5–6, 28, 48–67,
dialectal, 43, 47, 66
semiotics, 15
separative, 77, 83, 223, 233–235, 240–244, 247–250, 252–261, 265–269, 272–280, 296, 338, 342, 343, 354, 368, 371, 375, 379, 380, 382, 385, 391, 408, 434 n. 4, 6, 7, 435 n. 10, 11, 12, 436 n. 3
singular, *see* number
stative, 77, 83, 222, 240, 241, 248–252, 254, 272–280, 338, 339, 341–343, 354, 369, 379, 380, 382, 391, 434, 435
subjunctive, 62–63, 91, 128–130, 134, 135, 179, 188–192, 198–202, 203–204, 206, 286, 315, 335, 337–339, 341, 350, 354, 382, 431 n. 9
hortative, 198
orders, 103
see also indicative, mood, negation, non-perfective, present
substantivals, 85–86, 218–219, 221, 226–227, 233, 312, 315, 318–319, 391, 395, 432 n. 2 (ch. 16)
external case, 224, 233
substantival pronouns *see* pronouns
see also asymmetry
superlative, 29, 41, 145–155, 164, 338,

340–343, 354, 358, 379, 386, 428
suppletion, 86, 116, 120, 156, 157, 164, 199, 316, 399
system congruity, 14, 20

TAM *see* tense–aspect–mood
temporal, 39, 45, 83, 158, 188, 219, 230–232, 235, 237, 238, 259–264, 298, 306, 307, 310, 311, 433
tense, 79, 82–83, 126, 203–209, 208, 312, 313, 316, 317, 354, 368, 370, 375, 379, 388, 391, 398, 407
  non-remote, 189, 196, 334, 337, 376, 380–381
  see also tense–aspect–mood
tense–aspect–mood, 63, 68, 79, 82–85, 89, 91, 112, 115–120, 127, 138, 158, 188–194, 198, 202, 207, 314, 315, 394, 395, 399, 400, 433
thematic classes: *see* oikoclitic classes
transitivity, 57, 78–80, 211–216, 219, 234, 317, 334, 336–339, 341–343, 358, 360, 365, 366, 368, 370, 380–383, 397, 408, 432
transparency, 14, 18, 19, 25, 26, 34, 45, 46, 278, 283, 312, 362, 364, 368, 377, 382–385, 389, 394, 395, 404, 405, 434

uniform encoding, 18
universal, pronouns *see* pronouns

valency distinctions, 79, 211–217, 396–397
verbs:
  defective imperative, 126, 194, 197
  defective, 125
volition, 159, 160, 203, 207, 208, 335, 343, 354, 362, 382
vowel (system), 51, 71, 72, 80, 81, 104–105, 106, 108–109, 116–117, 122, 124, 131, 176, 182, 196, 197, 304, 314, 401, 427 n. 1, 428 n. 2 (ch. 10), 433 n. 6
  loss, 76
  oikoclitic vowel classes, 75
  vowel harmony, 135, 428 n. 11

word-forms, 18, 36, 114, 115, 218, 270, 279, 287, 288, 291, 294, 310, 311, 400, 402, 403

xenoclitic classes, 72, 73, 86, 92, 93, 98, 101, 143, 144, 227–229, 234, 324–333, 338, 339, 439
  xenoclitic adjectives, 95, 227, 332, 433
  xenoclitic verbs, 97, 134, 196
  see also adjectives, adverbs

yes/no particles, 156, 157

zero coding, 18, 221